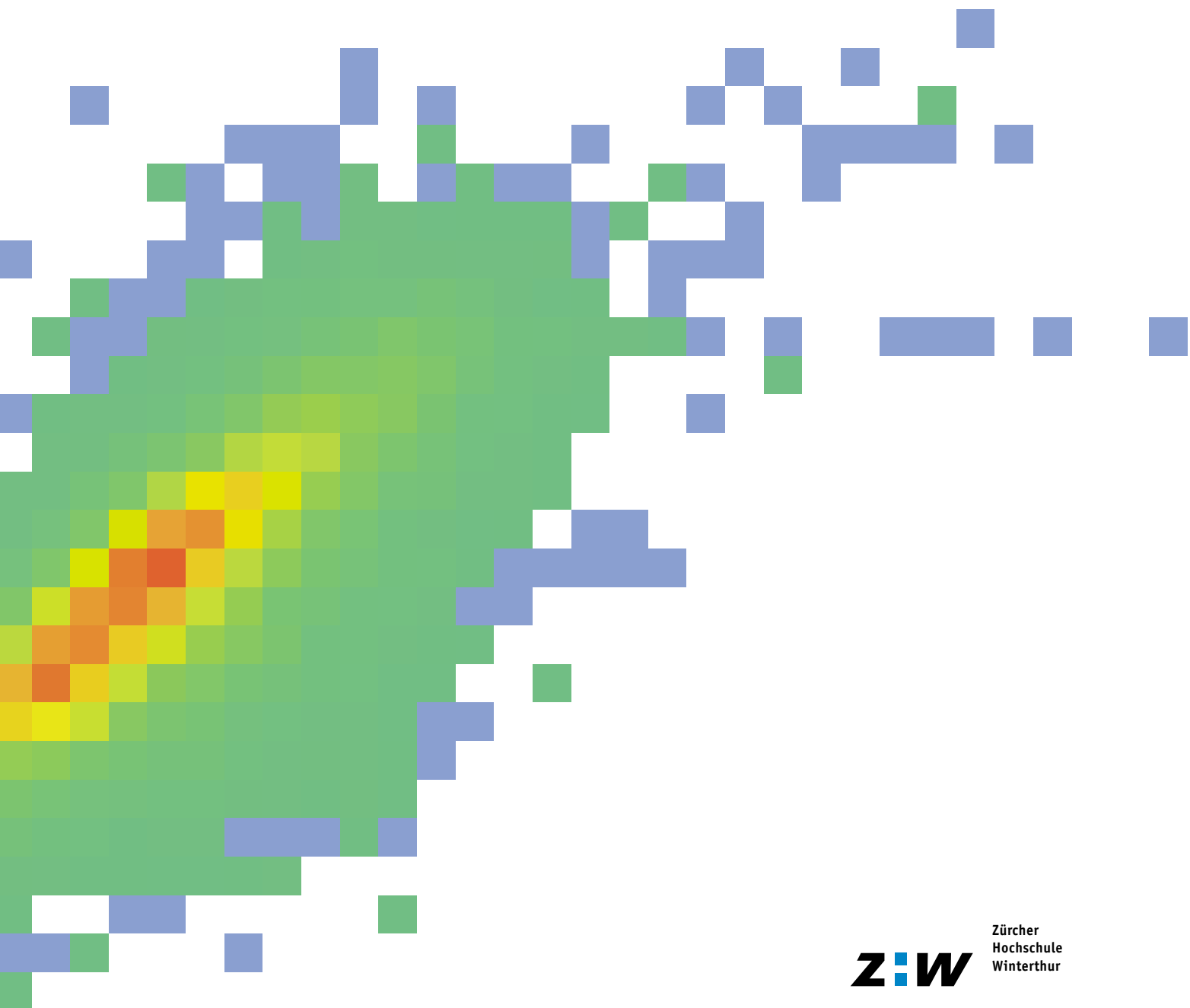




Plausibilisierung von Ozon, Stickoxiden und PM10

Statistische Methoden zur Effizienz- und
Qualitätssteigerung der Messdatenplausibilisierung



Z:W

Zürcher
Hochschule
Winterthur

Institut für Datenanalyse
und Prozessdesign

idp



Zusammenfassung	3
1. Einleitung	5
2. Grundsätzliche Betrachtungen zur Plausibilisierung	6
3. Statistische Umsetzung der Ideen	8
3.1 Transformation der Halbstundenwerte	9
3.2 Festlegung des Wertebereichs	9
3.3 Identifikation von Geräteausfall	10
3.4 Abweichungen vom erwarteten Verlauf (Sprünge)	10
4. Verwendete Daten	11
5. Plausibilisierung von Ozon und Stickoxiden	12
6. Plausibilisierung von PM10	23
7. Physikalisch-chemische Interpretationen	28
7.1 Charakterisierung der Standorte	28
7.2 Verteilung der Konzentration gasförmiger Schadstoffe	29
7.3 Verteilung der Partikelkonzentrationen	30
8. Literatur	31
9. Glossar	32

Impressum

Herausgeberin

Zentralschweizer Umweltschutzdirektionen (ZUDK) in Zusammenarbeit mit dem Kanton Aargau

Aktuelle Informationen sind im Internet unter www.in-luft.ch verfügbar

Verantwortliche Redaktion

René Locher und Andreas Ruckstuhl, Institut für Datenanalyse und Prozessdesign an der Zürcher Hochschule Winterthur, www.idp.zhwin.ch

Kontaktstellen

Umweltschutzämter der Kantone

Luzern:	Postfach, 6002 Luzern,	Telefon 041 228 60 60,	afu@lu.ch
Nidwalden:	Engelbergstr. 34, 6371 Stans	Telefon 041 618 75 04	afu@nw.ch
Obwalden:	Postfach 1661, 6061 Sarnen	Telefon 041 666 63 27	umwelt.luft@ow.ch
Schwyz:	Postfach 2162, 6431 Schwyz	Telefon 041 819 20 35	afu.di@sz.ch
Uri:	Klausenstrasse 4, 6460 Altdorf	Telefon 041 875 24 21	afu@ur.ch
Zug:	Postfach, 6301 Zug	Telefon 041 728 53 70	info.afu@bd.zg.ch
Aargau:	Buchenhof, 5001 Aarau	Telefon 062 835 33 60	umwelt.aargau@ag.ch

Gestaltung

Hilfiker AG, Werbeagentur BSW, Luzern



in-Luft benutzt zur Auswertung, Verteilung und Archivierung der Luftmessdaten die Software-Applikation eta® Power Desk. Zur Zeit beschränkt sich die Bereinigung der Daten auf eine Sichtung der Messwertverläufe und des Stationsjournals sowie der Interpretation von einzelnen Messsystemmeldungen in Form von sogenannten Status-Codes. Dies bedeutet, dass spezialisierte Personen mit höchster Konzentration u. a. die Messwertverläufe am Bildschirm visualisieren müssen, um bei Auffälligkeiten die entsprechenden Zeitabschnitte genauer zu untersuchen. Oft handelt es sich bei den der Fachperson «ins Auge gefallenen» Werten um besondere lufthygienische Ereignisse, welche nicht auf einen Defekt im Mess- oder Datenübertragungssystem zurückgeführt werden können. Die Daten bleiben in diesem Fall konsequenterweise unverändert in der Datenbank. Wenn die auffälligen Werte mit grösster Wahrscheinlichkeit auf eine Fehlfunktion des Systems zurückzuführen sind, werden die entsprechenden Daten in der Datenbank als ungültig markiert und von weiteren Berechnungen und Publikationen ausgeschlossen.

Dieses Vorgehen ist mit zwei Unzulänglichkeiten behaftet: Zum einen durchlaufen die Daten, welche aus Aktualitätsgründen automatisch, z. B. via Internet, publiziert werden, erst nachträglich eine Plausibilisierung. Zum andern ist die visuelle Überprüfung *aller* Daten aufwendig und für die durchführende Person sehr ermüdend.

In dieser Arbeit wurden Vorschläge zum automatischen Auffinden von auffälligen Ozon-, Stickoxid- und PM10-Werten erarbeitet, welche einfach in eta zu implementieren sind. Auffällig werden in diesem Zusammenhang Werte bezeichnet, deren Eigenschaften stark von den erwarteten abweichen und somit potentiell fehlerbehaftet sind. Die Implementation dieses Verfahrens erlaubt der Fachperson, sich bei der Plausibilisierung der Luftmessdaten in Zukunft vor allem auf diese auffälligen Stellen zu konzentrieren. Damit kann sie gleichzeitig ihre Effizienz steigern und die Gefahr verringern, dass fehlerbehaftete Stellen in den Daten übersehen werden.

Das vorgeschlagene Verfahren wurde an Datensätzen von vier in ihrer lokalen Situation sehr unterschiedlichen Standorten überprüft:

- Lungern-Schönbühl (nur O₃-Messungen, Hintergrundstandort auf 1550 m ü. Meer)
- Schwyz (Regionalzentrum)
- Luzern (Stadtstandort ohne direkte Verkehrsbeeinflussung)
- Erstfeld (Standort unmittelbar neben der Gotthard-Autobahn)

Die verwendeten Datensätze waren vorgängig durch in-Luft mit dem üblichen Verfahren manuell bereinigt worden und sollten somit keine Daten von schlechter Qualität mehr enthalten. Die Parameter der hier vorgestellten Verfahren wurden so eingestellt, dass unabhängig vom Standorttyp pro Schadstoff und Jahr im bereinigten Datensatz 5 bis 20 Stellen als auffällig markiert werden. Die auffälligen Stellen waren dabei relativ gleichmässig über das Jahr verteilt. Je nach Schadstoff wurden zwei (NO, NO₂, O₃) bzw. drei (NO_x, O_x, PM10) Regeln aufgestellt, welche von unauffälligen Werten gleichzeitig eingehalten werden müssen. Mit Ausnahme des zulässigen Wertebereichs konnten alle Regeln standortunabhängig gestaltet werden. Auf Grund der absichtlich sehr eng gesetzten zulässigen Wertebereiche könnten in Jahren mit aussergewöhnlichen Wetterbedingungen (extrem sonnenreicher und heisser Sommer oder ein Winter mit lang anhaltenden Inversionen) zu gewissen Zeiten gehäuft auffällige Werte auftreten, welche damit auf die aussergewöhnliche Situation des entsprechenden Jahres hinweisen würden.



Als Nebenprodukt der Entwicklung dieser Plausibilisierungsverfahren wurden auch interessante lufthygienische Erkenntnisse gewonnen. Es zeigte sich z. B., dass die O_x -Werte zeitlich und örtlich viel glatter¹ sind als die entsprechenden O_3 -Werte. Das Verhalten für die PM_{10} -Werte liegt zwischen demjenigen von O_x und O_3 . Grössere O_x -Konzentrations sprünge in Tälern des Alpengebiets können auf Föhnereignisse hinweisen: In der Hälfte der O_x -Sprünge (9 von 17) war dies z. B. in Erstfeld der Fall. Bei steigenden Temperaturen nahm die Luftfeuchtigkeit sprunghaft ab und gleichzeitig die Windgeschwindigkeit deutlich zu. In vier weiteren Fällen konnten die aussergewöhnlich hohen Sprünge ebenfalls mit zunehmendem Wind (ohne Änderung der Temperatur und Luftfeuchte) erklärt werden.

¹ d. h. die Werte ändern sich von Zeitpunkt zu Zeitpunkt bzw. von einem Ort zum andern nur langsam.

1. Einleitung



in-Luft benutzt zur Auswertung, Verteilung und Archivierung der Luftmessdaten die Software-Applikation eta® Power Desk, im Folgenden kurz eta genannt. Dieses von Zühlke Engineering AG entwickelte Softwarepaket erlaubt

- das automatische Importieren der Messdaten von den einzelnen Messstationen,
- das automatische Berechnen von in der Luftreinhalteverordnung verlangten statistischen Kenngrößen,
- das elektronische Versenden von individuell zusammengestellten Kenngrößen an ausgewählte Kunden sowie
- die automatische Publikation von Immissionswerten via Webseiten und SMS.

Zur Zeit beschränkt sich die Bereinigung der Daten auf eine Sichtung der Messwertverläufe und des Stationsjournals sowie die Interpretation von einzelnen Messsystemmeldungen in Form von sogenannten Status-Codes. Dies bedeutet, dass spezialisierte Personen mit höchster Konzentration u. a. die Messwertverläufe am Bildschirm visualisieren müssen, um bei Auffälligkeiten die entsprechenden Zeitabschnitte genauer zu untersuchen. Oft handelt es sich bei den der Fachperson «ins Auge gefallen» Werten um besondere lufthygienische Ereignisse, welche nicht auf einen Defekt im Mess- oder Datenübertragungssystem zurückgeführt werden können. Die Daten bleiben in diesem Fall konsequenterweise uneditiert in der Datenbank. Falls das Ereignis für das allgemeine Publikum von Interesse ist, werden diese auffälligen Werte als Datengrundlage für einen Sonderbericht oder eine Pressemitteilung verwendet. Wenn die auffälligen Werte mit grösster Wahrscheinlichkeit auf eine Fehlfunktion des Systems zurückzuführen sind, werden die entsprechenden Daten in der Datenbank als ungültig markiert und von weiteren Berechnungen und Publikationen ausgeschlossen.

Dieses Vorgehen ist mit zwei Unzulänglichkeiten behaftet: Zum einen durchlaufen die Daten, welche aus Aktualitätsgründen automatisch publiziert werden, erst nachträglich eine Plausibilisierung. Zum andern ist die visuelle Überprüfung *aller* Daten aufwendig und für die durchführende Person sehr ermüdend.

Ziel dieser Arbeit war deshalb das Erarbeiten von Erkenntnissen, welche in Zukunft die automatische Identifikation auffälliger Werte erlaubt. Eine besondere Herausforderung war dabei die notwendige präzise Definition, was unter auffälligen Werten zu verstehen ist. Die Definition sollte so erfolgen, dass die auffälligen Werte möglichst alle Daten mit groben Fehlern und gleichzeitig möglichst wenig plausible Daten umfassen. Die vorgeschlagenen Rechenverfahren sollten dabei möglichst einfach interpretierbar und in die Software implementierbar sein.

Der vorliegende Bericht richtet sich an lufthygienisches Fachpersonal. Fachbegriffe aus der Statistik und spezielle Definitionen, welche in diesem Bericht verwendet werden, sind im Glossar, Kap. 9 zusammengefasst. Um die Programmierung der Analysen zu vereinfachen und die Dokumentation des Berichts einheitlich zu gestalten, wurden alle Messzeiten in diesem Bericht (im Gegensatz zum heute noch üblichen schweizerischen Lufthygienestandard) als **Startzeiten** dokumentiert. Bei von-bis-Angaben wird die Startzeit für den ersten Wert und die Endezeit des letzten Wertes angegeben.

2. Grundsätzliche Betrachtungen zur Plausibilisierung



Wenn eine Fachperson manuell Daten plausibilisiert, so verlässt sich diese bewusst und auch unbewusst auf eine Vielzahl von Regeln, welche sie sich im Verlauf der beruflichen Erfahrung angeeignet hat. Die Regeln lassen sich grob in vier Kategorien einteilen.

- **Wertebereich:** Die überprüften Werte müssen innerhalb eines bestimmten Bereiches liegen. Allenfalls gelten je nach Bedingungen andere Wertebereiche. Z. B. muss die Globalstrahlung während der Nacht innerhalb der Messgenauigkeit 0 W/m^2 betragen, während sie unter Tags zwischen 0 und 1200 W/m^2 beträgt. Bei einer genaueren Betrachtung sollte die obere Grenze sogar je nach Stunde des Tages und Tag des Jahres sowie dem Breitengrad der Messstation und deren Horizontüberhöhung² variiert werden.
- **Folgen (fast) identischer Werte (Geräteausfall):** Diverse Ursachen können zu Folgen mit vielen (fast) identischen Werten führen wie z. B. der Verlust des Kontakts zwischen dem ein Analogsignal liefernden Kabel und dem analog-digital-Konverter in der Datenerfassung oder ein Programmierfehler im Messgerät, der Datenerfassung oder -übertragung³.
- **Plötzliche Abweichung vom plausiblen Verlauf der Zeitreihen (Detektion von Sprüngen):** Bei defekten Geräten, Wackelkontakt oder elektromagnetischen Störfeldern können einzelne positive oder negative Spikes oder Stufen auftreten, oder das Signal ist möglicherweise generell stark verrauscht. Analoge Fehler können auch nach Fehlmanipulationen an Geräten auftreten wie dem unbeabsichtigten Verstellen des Empfindlichkeitsbereichs des Messgerätes.
- **Abhängigkeiten zwischen verschiedenen Messgrößen:** Je nach Messgrösse erleichtern weitere unabhängig gemessene Größen die Plausibilisierung (vgl. Kap. 5).

Als *ungültige* Daten werden in diesem Bericht wie auch in der eta Werte bezeichnet, welche mit grosser Sicherheit nicht den zur Messzeit herrschenden «wahren» Wert wiedergeben.

Als *auffällig* werden Daten definiert, wenn sie mindestens einer der obigen Regeln nicht genügen. Auffällige Werte sind nicht à priori auch ungültige Werte und umgekehrt. Die Begriffe «ungültig» und «auffällig» sind somit nicht deckungsgleich.

Ein Beispiel zur Illustration: Wird zu einem Zeitpunkt t eine Kalibration mit O_3 -Kalibriergas durchgeführt, dessen Konzentration sich im Bereich der in der Aussenluft erwarteten Konzentration bewegt, so ist der gemessene Wert möglicherweise nicht auffällig, wenn nur jeweils eine Wertebereichsregel, eine Regel für den zeitlichen Verlauf und eine Regel für lange Serien gleichbleibender O_3 -Halbstundenwerte aufgestellt wurde. Trotzdem gibt der Wert höchstens rein zufällig die richtige Aussenluftkonzentration wieder! Der Wert ist also unauffällig aber ungültig. Der Unterschied zwischen ungültig und auffällig kann mit der Tatsache begründet werden, dass die Regeln praktisch nie das Verhalten des Messprozesses vollständig abbilden. In unserem Beispiel fehlte eine Regel zum Zustand des Messgerätes, ob es sich nämlich gerade im Kalibrationszustand befindet oder nicht.

Umgekehrt können wegen unvollständiger Abbildung der Realität auffällige Werte trotzdem gültig sein. Ein Beispiel dazu finden Sie für den NO -Verlauf in Abb. 5-5: Bei der Detektion von aussergewöhnlich hohen Sprüngen, d. h. von Sprüngen $> 100 \text{ ppb(!)}$ würden häufig Fehlalarme ausgelöst.

² vgl. z. B. verspätetes Aufgehen der Sonne hinter einem Gebirge

³ Wenn der Programmierfehler genügend «raffiniert» ist, so wird dieser nur sehr sporadisch auftreten, nämlich dann wenn gerade die entsprechenden Bedingungen eingetreten sind.

2. Grundsätzliche Betrachtungen zur Plausibilisierung



Als *kritisch* werden hier und in der eta Werte bezeichnet, welche auffällig sind und durch die Fachperson «gefühlsmässig» als sehr zweifelhaft beurteilt werden. Trotzdem findet die Fachperson kein hartes objektives Argument, um den betreffenden Wert definitiv als ungültig zu erklären. Die Fachperson überlässt also gewissermassen den nachfolgenden Stellen die Wahl, ob die entsprechenden Werte in den Auswertungen berücksichtigt werden sollen oder nicht.

In diesem Zusammenhang ist noch eine weitere Bemerkung wichtig. Wenn Zeitreihen verdichtet werden, z. B. von Halbstundenmittel auf Tagesmittel, so wird die Qualität der Tagesmittel nicht nur durch aussergewöhnliche sondern auch durch weggelassene Werte in praktisch jedem Fall negativ beeinflusst. Durch geeignete Interpolationen könnte der negative Aspekt der Elimination von ungültigen Werten gemildert werden. Dies ist jedoch nicht Gegenstand dieser Arbeit.

Aus den obigen Betrachtungen können folgende Schlüsse gezogen werden:

- Automatische Vorplausibilisierungen können zwar via Suche nach auffälligen Werten durchgeführt werden. Die definitive Ungültig-Setzung sollte jedoch einer Fachperson überlassen werden.
- Eine automatische Plausibilisierung kann nie alle ungültigen Werte herausfiltern – eine Fachperson auch nicht. In jeder Hinsicht völlig unauffällige ungültige Werte wird auch eine Fachperson nicht finden. Die Fachperson findet nur dann zusätzliche ungültige Werte, wenn gewisse Regeln schlecht oder gar nicht im Programm abgebildet wurden.
- Das Programm macht im Gegensatz zum Menschen keine Flüchtigkeitsfehler. Bei gut programmierten Regeln hilft die automatische Vorplausibilisierung, Flüchtigkeitsfehler bei der manuellen Plausibilisierung zu vermindern. Die Fachperson muss nicht mehr alle Werte kontrollieren sondern kann sich auf eine begrenzte Auswahl von potentiell ungültigen Werten konzentrieren.
- Gleichzeitig fördert die automatische Vorplausibilisierung die Standardisierung der Datenqualität, indem gleiche Merkmale im Datenverlauf bei der Vorplausibilisierung immer gleich behandelt werden. Weiter zwingt die Programmierung der automatischen Vorplausibilisierung die Lufthygienefachleute dazu, sich vorgängig detaillierte Gedanken über allgemein gültige Regeln bezüglich der Datenqualität zu machen und somit das intuitive, latent vorhandene Wissen zu systematisieren und zu dokumentieren.

3. Statistische Umsetzung der Ideen



Gewisse Fehler oder Unzulänglichkeiten, welche während eines Messprozesses auftreten, sind zwar grundsätzlich bekannt, können aber im Rahmen der finanziellen und personellen Möglichkeiten nicht eliminiert und müssen deshalb in Kauf genommen werden. Häufig behilft man sich in einem solchen Fall durch die Definition einer sogenannten Referenzmethode und stellt damit sicher, dass wenigstens die systematischen Abweichungen vom «wahren» Wert bei all denjenigen Verfahren gleich sind, welche äquivalent zu dieser Referenzmethode sind.

Häufig auftauchende Fehler werden wenn immer möglich durch technische Massnahmen abgefangen, so dass in der Regel nur Fehler mit kleiner Auftretenswahrscheinlichkeit übrigbleiben. Entsprechend vielfältig und schlecht definiert sind dann diese noch übrig bleibenden Fehler, welche wenn möglich erkannt werden sollten.

Als Ausgangspunkt für die hier beschriebenen Arbeiten wurden deshalb die real gemessenen Werte genommen, welche von Hand bereits eingehend plausibilisiert worden sind. An diese Daten wurde dann ein möglichst enges Korsett von Regeln im Sinne der vier möglichen Fehlerkategorien angelegt, wie sie in Kap. 2 beschrieben werden. Die Regeln wurden so eng definiert, dass pro Jahr und Standort insgesamt bis zu zwei Dutzend Stellen als auffällig markiert wurden. Dies entsprach rund 1 Promille der zu überprüfenden Daten. Die auffälligen Stellen waren bis auf wenige Ausnahmen vom lufthygienischen Standpunkt aus immer noch plausibel⁴. Die Idee dahinter basierte auf den Überlegungen, wie sie in Kap. 2 bereits dargelegt wurden: Ob Mensch oder Maschine, es können nur ungültige Werte gefunden werden, welche besonders auffällig sind. Um möglichst viele der ungültigen Werte zu finden, wird in Kauf genommen, dass auch einige wenige Werte als auffällig angezeigt werden, welche aus lufthygienischer Sicht in Ordnung sind. Die Menge muss allerdings klein gehalten werden, da sonst die Aufmerksamkeit der plausibilisierenden Fachperson bei der Überprüfung über Gebühr nachlässt.

Ferner muss beachtet werden, dass die hier vorgeschlagenen Plausibilisierungen vor jeglicher Korrektur vorgenommen werden sollten. Die Korrektur gemäss den Messempfehlungen des BUWAL sieht z. B. vor, dass leicht negative Konzentrationen, welche physikalisch keinen Sinn machen, auf 0 gesetzt werden [1]. Mit dieser Korrektur könnten somit u. U. künstlich längere Folgen von Nullwerten entstehen, wenn die gemessenen Werte im Bereich der Nachweisgrenze des Messgerätes lagen. Aus den umgekehrten Gründen muss die Plausibilisierung auch vor der Kalibrationskorrektur bzw. vor der Umrechnung der TEOM-PM10-Werte in HVS-PM10-Werte vorgenommen werden. Bei diesen beiden Korrekturen könnten potentiell identische Werte, welche auf einen Geräteausfall hinweisen, durch die Korrekturen in Folgen unterschiedlicher Werte verwandelt werden, welche anschliessend nicht mehr mit unserer Methode detektiert werden könnten.

⁴ Es wurden ja bereits manuell plausibilisierte Daten verwendet!

3. Statistische Umsetzung der Ideen



3.1 Transformation der Halbstundenwerte

In der Statistik werden häufig vor der eigentlichen Analyse Transformationen der ursprünglichen Grössen durchgeführt. Sie bezwecken die Überführung der Daten bzw. – nach Anwendung eines Modells – der Residuen in eine Verteilung, welche möglichst nahe derjenigen einer Normal-Verteilung mit konstanter Varianz entspricht.

Bei der Analyse von Konzentrationen macht man oft die Erfahrung, dass die logarithmierten Werte eine Normal-Verteilung aufweisen: Man spricht in diesem Fall von lognormal-verteilten Konzentrationen. Diese Eigenschaft weist darauf hin, dass viele (kleine) Effekte, welche die Konzentrationen beeinflussen, eher multiplikativer als additiver Natur sind.

Als weiteres heuristisches Argument für die Nützlichkeit von log-Transformationen bei Konzentrationen kann die Tatsache herangezogen werden, dass die bei statistischen Verfahren häufig vorausgesetzte Normal-Verteilung per Definitionem immer auch Werte im negativen Bereich aufweist⁵. Modelle, welche *nicht*-transformierte Konzentrationen beschreiben, können also auch negative Konzentrationen als Resultate ausweisen. Dies ist bekanntlich physikalisch unsinnig. Werden die Konzentrationen vor der Anwendung des Modells logarithmiert, so verteilen sich die Werte im ganzen reellen Zahlenbereich. Werte zwischen 0 und 1 im untransformierten Bereich entsprechen dabei den negativen Werten im log-transformierten Bereich. Allerdings entspricht eine Konzentration von 0 nach der Logarithmustransformation $-\infty$ und darf somit (wie auch negative Werte) im zu transformierenden Datensatz nicht vorliegen. Die Konzentration 0 existiert für reell existierende Stoffe nicht. In der Praxis können diese aber wegen der beschränkten Messgenauigkeit der Geräte genau so auftauchen wie negative Werte und müssen als Artefakte vor der Transformation entsprechend «vorbehandelt» werden.

Handelt es sich bei den interessierenden Grössen um Anzahlen pro Zeit- oder Volumeneinheit, so sind deren Werte in der Regel poisson-verteilt. Die Varianz ist somit proportional zur erwarteten Anzahl. Die Wurzel-Transformierte einer solchen Grösse ist wiederum eine Verteilung mit konstanter Varianz und gleicht sich mit zunehmender Anzahl immer mehr einer Normal-Verteilung an.

Falls z. B. ein bestimmter physikalisch-chemischer Effekt die Messwerte stark beeinflusst, ist es manchmal sinnvoller eine Transformation vorzunehmen, welche mehrere gemessene Grössen in eine neue, abstraktere überführt, und die statistische Analyse an der transformierten Grösse durchzuführen.

3.2 Festlegung des Wertebereichs

Für die Festsetzung der Wertebereiche wurde von den langjährigen Erfahrungswerten von in-Luft ausgegangen und die Maxima allenfalls nach unten korrigiert, wenn während des Stichjahres die Maximalwerte massiv unter dem erlaubten Maximum lagen. Die unteren Grenzwerte für Konzentrationen entsprechen grob der mit -1 multiplizierten Messunsicherheit der entsprechenden Messgeräte im tiefen Konzentrationsbereich. Mit den zugelassenen negativen Konzentrationen, welche an sich physikalisch keinen Sinn machen, wird darauf Rücksicht genommen, dass die Messprozedur auch bei korrekter Anwendung und tadellosem Zustand der Messgeräte bei sehr tiefen Konzentrationen (leicht) negative Werte liefern kann.

Bei diesem Vorgehen wurde also bewusst in Kauf genommen, dass vereinzelt auch gültige Werte ausserhalb dieses Bereiches zu liegen kommen. Da diese Werte weit entfernt vom Grossteil der übrigen Werte liegen, sind sie jedoch von allgemeinem Interesse. Die auffälligen Werte können sich auch in gewissen ausserordentlichen Jahren massieren während üblicherweise nur sehr wenige bis gar keine Werte ausserhalb des Wertebereiches gefunden werden.

⁵ Bei einer sehr engen Normalverteilung mit einem Erwartungswert, welcher weit von 0 entfernt im positiven Bereich liegt, ist die Wahrscheinlichkeit von negativen Werten allerdings sehr klein.

3. Statistische Umsetzung der Ideen



3.3 Identifikation von Geräteausfall

Die Identifikation von Folgen (fast) identischer Werte wurde durch Überprüfung der gleitenden Standardabweichung über jeweils n nachfolgende Werte realisiert: Wenn überhaupt keine Änderung des Wertes in einer Folge von n Werten auftritt, so beträgt die Standardabweichung s exakt 0. Falls auch Folgen von n fast identischen Werten detektiert werden sollen, so sind je nach Grösse von n und je nach erlaubter Grösse der Wertefluktuations auch etwas grössere Werte von s erlaubt. In diesem Zusammenhang ist vor allem auf die Anzahl der für die Plausibilisierung relevanten Stellen der Daten zu achten. Obwohl für die Kommunikation gegen aussen nur signifikante Stellen angegeben werden sollten, sind hier auch die nicht signifikanten von Interesse! Mit einer zunehmenden Anzahl von Stellen sinkt die Wahrscheinlichkeit, dass Werte rein zufällig, d. h. ohne Defekt im System, exakt identisch sind. Mit zunehmender Anzahl Stellen kann damit die Trennschärfe der Unterscheidung erhöht werden zwischen rein zufälligen (fast) identischen Werten und solchen, welche durch einen Defekt verursacht sind.

Als Grenzwerte für s zeigten sich für die Lufthygienedaten 0.005 und 0.05 Einheiten von Bedeutung (vgl. Kap. 5 und 6). Ein Wert von s von unter 0.005 bedeutet drei identische Werte, sofern die Auflösung der Werte maximal 0.01 Einheiten beträgt⁶. Ein Wert von s von unter 0.05 bedeutet 4 identische Werte, sofern die Auflösung der Werte maximal 0.1 Einheiten beträgt.

3.4 Abweichungen vom erwarteten Verlauf (Sprünge)

Die Überprüfung auf Abweichungen vom erwarteten Verlauf von Zeitreihen setzt bei Aussenluftmessungen voraus, dass die Zeitreihe glatt ist, d. h. dass die Werte der Zeitreihe sich von Zeitpunkt zu Zeitpunkt nur langsam ändern. Bei genügend glatten Zeitreihen reicht die Analyse einfacher Differenzen von jeweils zwei sich unmittelbar folgenden Werten.

An Stelle der einfachen Differenzen kann auch eine lokale robuste⁷ Regression verwendet werden, um Abweichungen vom erwarteten Verlauf festzustellen. Bei der lokalen Regression⁸ werden die Messpunkte unmittelbar vor dem zu überprüfenden Zeitpunkt t für eine Regression verwendet, um den Wert zum Zeitpunkt t vorherzusagen. Dabei gehen die Punkte mit zunehmender zeitlicher Distanz zu t mit kleinerem Gewicht in die Regression ein. Als Residuum wird die Differenz zwischen dem vorhergesagten und dem gemessenen Wert bezeichnet. Je grösser das Residuum ausfällt, um so weiter weicht der gemessene Punkt vom erwarteten Wert ab. Das Verfahren ist programmiertechnisch viel aufwendiger als die einfache Differenzenbildung, kann aber bei stark verrauschten Daten oder bei Daten mit extremer Dynamik zu verlässlicheren Ergebnissen führen. In dieser Arbeit zeigte sich, dass die lokale robuste Regression, welche die nächsten 8 Werte vor dem zu beurteilenden Punkt berücksichtigte, leicht bessere Resultate erzeugte als die einfache Differenzenbildung. Angesichts des viel kleineren Aufwandes für die einfache Differenzenbildung wurde der Vorteil aber als zu klein erachtet. Im Weiteren werden deshalb nur die Resultate aus der Differenzenbildung beschrieben.

⁶ oder, bei mehr Nachkommastellen, zwei identische Werte plus ein dritter, welcher weniger als 0.008 Einheiten von den andern beiden abweicht.

⁷ vgl. robuste Schätzung in Glossar

⁸ vgl. Regression, robust in Glossar

4. Verwendete Daten



Um die Allgemeingültigkeit der vorgeschlagenen Plausibilisierungsalgorithmen abschätzen zu können und deren Parameter für beliebige Messstationen vorschlagen zu können, wurden die Daten (vgl. Tab. 4-1) von vier lufthygienisch sehr unterschiedlichen Standorttypen untersucht (vgl. Tab. 4-2). Dabei wurden, mit Ausnahme von Lungern, jeweils 19 Monate Messdaten (1.6.00 bis 31.1.02) in die Analysen einbezogen. Wenn nicht anders vermerkt, wurden jedoch den Übersichtstabellen und Grafiken dieses Berichts nur die Daten vom 1.2.01 bis 31.1.02 zu Grunde gelegt.

Die Station Lungern wurde als Hintergrundstandort ausgewählt (Messungen von Ozon, jeweils von April bis September). Schwyz repräsentiert ein Regionalzentrum mit mittlerer Belastung. Die Station Luzern wurde als typischer Stadtstandort in die Untersuchung einbezogen. Als Verkehrsstandort wurde schliesslich Erstfeld ausgewählt, welcher sich unmittelbar neben der Gotthard-Autobahn befindet.

Abkürzung	Erklärung
O ₃ , NO _x , NO	erfasste gasförmige Schadstoffe in ppb
NO ₂	NO _x – NO
O _x	O ₃ + NO ₂ = O ₃ + NO _x – NO
PM10	PM10, gemessen mit bei 30°C thermostatisiertem TEOM, in µg/m ³
T	Lufttemperatur in °C
p	Luftdruck in hPa
T _d	Taupunkt in °C
p _{H2O}	Partialdruck des Wassers in hPa
H _{rel}	relative Luftfeuchte
WV _v	vektoriell berechnete Windgeschwindigkeit in m/s
WV _s	skalar berechnete Windgeschwindigkeit in m/s
WD	Windrichtung = Winkel des über 30 min integrierten Windweges: 0° = 360° = Wind aus Norden, 90° = Wind aus Osten.
T _i	Innentemperatur

Tab. 4-1: Abkürzungen der analysierten Messgrössen

Standort (Abkürzung)	erfasste Messgrössen
Lungern-Schönbüel, 1550 m ü. M	O ₃
Schwyz, Rubiswilstrasse, 470 m ü. M	O ₃ , NO _x , NO, PM10, Str _{Glo} , T, p, T _d , p _{H2O} , H _{rel} , WV _v , WV _s , WD, T _i
Luzern, Museggstrasse 7a, 460 m ü. M	O ₃ , NO _x , NO, PM10, T, p, T _d , p _{H2O} , H _{rel} , T _i
Erstfeld, 460 m ü. M	O ₃ , NO _x , NO, PM10, Str _{Glo} , T, p, T _d , p _{H2O} , H _{rel} , WV _v , WV _s , WD, T _i

Tab. 4-2: Messstandorte und erfasste Messgrössen

5. Plausibilisierung von Ozon und Stickoxiden



Die Analyse der Ozon- und Stickoxiddaten zeigte einen zeitlich wenig glatten Verlauf. Dies galt besonders für Messstellen, welche sich in der Nähe von grossen Stickoxid-Quellen befanden. Dies hat zur Folge, dass die direkte Anwendung der Methode der einfachen Differenzen auf diese Schadstoffe versagt. Auf Grund der folgenden lufthygienischen Überlegung wurde schliesslich die entscheidende Transformation gefunden, welche die Konzentrationsverläufe in eine zeitlich glatte Funktion bringt:

- Der Aufbau von Ozon erfolgt über mehrere Stunden, während dessen Abbau via «Titration» mit NO zu NO₂ und Sauerstoff innerhalb von Sekunden erfolgt.
- Pro (durch «NO-Titration») abgebauten Ozonmolekül entsteht genau ein NO₂-Molekül.
- Die übrigen relevanten Abbaureaktionen und Verdünnungseffekte von Ozon sind im Vergleich zur «NO-Titration» langsam. Deren Geschwindigkeit liegt in der gleichen Grössenordnung wie diejenige des Ozonaufbaus durch Photooxidation oder sind sogar langsamer.

Geschwindigkeit des Ozonauf- und -abbaus:

innerhalb von Sekunden: $O_3 + NO \longrightarrow O_2 + NO_2$

innerhalb von Stunden: O₃-Aufbau durch Photooxidation
O₃-Abbau durch trockene Deposition etc.

Die unterschiedlichen Reaktionsgeschwindigkeiten für Ozonaufbau und -abbau erklären sehr gut die Tatsache, dass der Ozonverlauf weit abseits der Stickoxidquellen zeitlich viel glatter verläuft als in deren Nähe. Aus analogen Überlegungen sollte deshalb O_x, die Summe aus O₃ und NO₂, zeitlich glatter verlaufen als die einzelnen Summanden für sich alleine genommen. Ferner sollte der zeitliche Verlauf von O_x sich auch an unterschiedlichen Orten viel ähnlicher verhalten als der O₃-Verlauf.

Diese theoretischen Betrachtungen lassen sich sehr schön mit entsprechenden Grafiken belegen. In Abb. 5-1 und Abb. 5-2 sind die Lag-Plots der logarithmierten⁹ Konzentrationen von O_x bzw. O₃ für die Messstation in Schwyz aufgetragen. Lag-Plots werden zur Analyse der Korrelationsstruktur von Zeitreihen verwendet: In Richtung der x-Achse werden die Werte der zu analysierenden Zeitreihe aufgetragen. In Richtung der y-Achse werden die zu den x-Werten um 1, 2, 3 usw. Zeiteinheiten¹⁰ verschobenen x-Werte aufgetragen. Die Grafik ist ganz analog zu einem normalen Streudiagramm zu interpretieren. Je stärker die Werte mit ihren Vorgängern (linear) korrelieren, umso enger streuen die Punkte im Lag-Plot um eine Gerade. Falls wir weisses Rauschen sehen, d.h. falls die Punkte zufällig in der x-y-Ebene des Lag-Plots verteilt sind, ist keine Korrelation vorhanden. In der x-y-Ebene kreisförmig verteilte Punkte sind ein Beispiel für solche unkorrelierte x-y-Daten.

Da hier sehr viele Punkte analysiert wurden – ein Jahr besteht aus 17 520 Halbstunden – wurde der normale (einfarbige) Lagplot modifiziert: Es wurden statt einzelner Punkte in der x-y-Ebene Punktedichten dargestellt. Die Skala am Rand der Grafik gibt die Anzahl Punkte pro Farbquadrat an.

Beim Vergleich der Lag-Plots für O₃ und O_x ist die höhere zeitliche Korrelation bei O_x deutlich sichtbar. Während bei O_x die Verteilung der Punkte auch noch bei einer zeitlichen Verschiebung von 4 h (lag = 8) zigarrenförmig ist, ist die Verteilung der O₃-Werte bei der gleichen zeitlichen Verschiebung praktisch gleich breit wie lang. Die Verteilung nähert sich der Kreissymmetrie und wird auch flacher (fast keine Punkte im roten und gelben Bereich mehr). Während die Korrelation der O_x-Konzentrationen im ganzen Konzentrationsbereich gleich hoch ist, sind bei den O₃-Werten im höheren Konzentrationsbereich, d.h. während den extremen Sommersmog-Phasen, deutlich höhere Korrelationen zu beobachten als im übrigen Bereich.

⁹ In dieser Arbeit wurde immer der zweier-Logarithmus verwendet, so dass die Skalen rechnerisch einfach zu interpretieren sind: Eine Differenz von 1 auf der entsprechenden Achse entspricht jeweils zwei um einen Faktor 2 verschiedenen Werten.

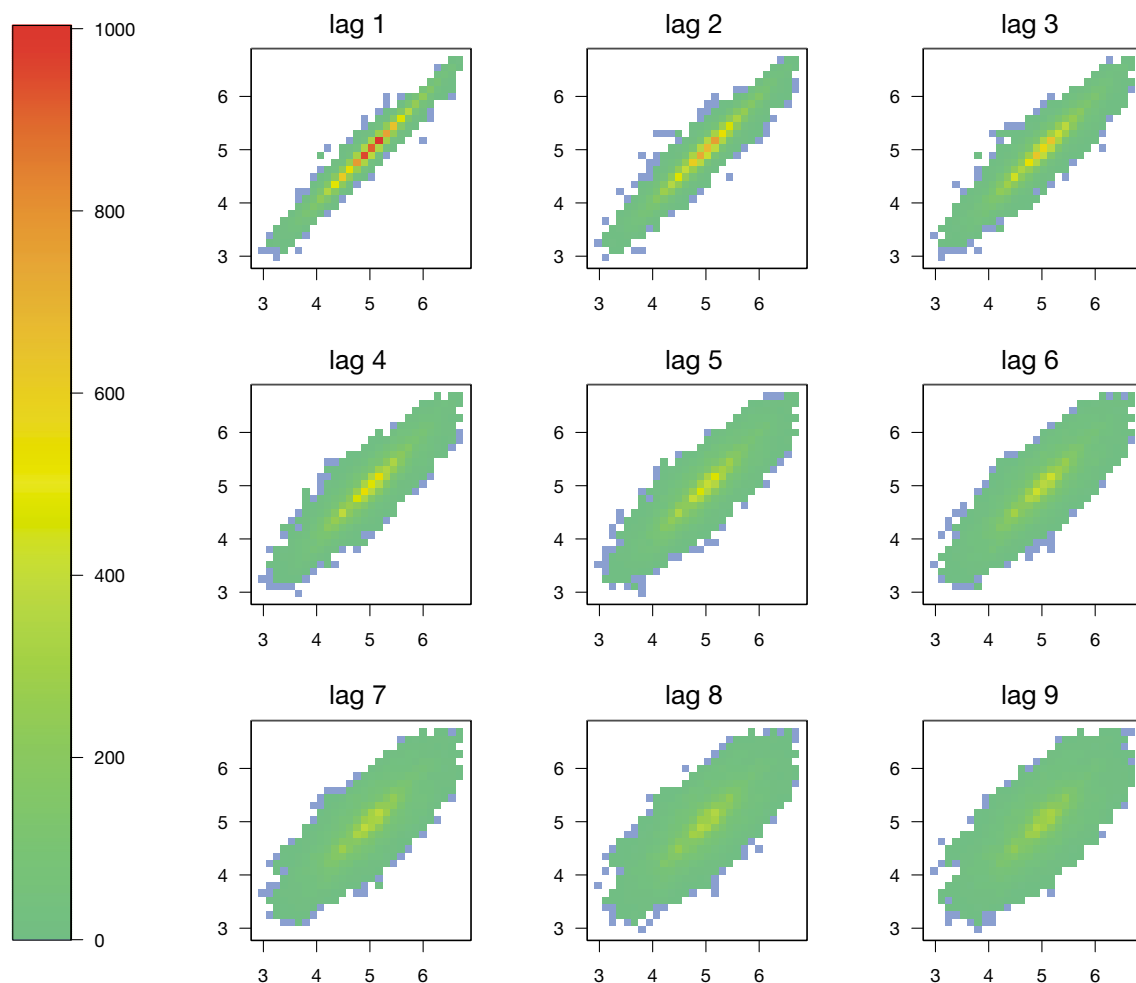
¹⁰ Eine Zeiteinheit beträgt in unserem Fall 30 min.

5. Plausibilisierung von Ozon und Stickoxiden



Abb. 5-1: Modifizierter Lag-Plots von $\log_2(O_3)$ für Schwyz

Die Skala am Rand der Grafik gibt die Anzahl Punkte pro Farbquadrat an. Einheiten der x- und y-Achsen $\log_2(\text{ppb})$.

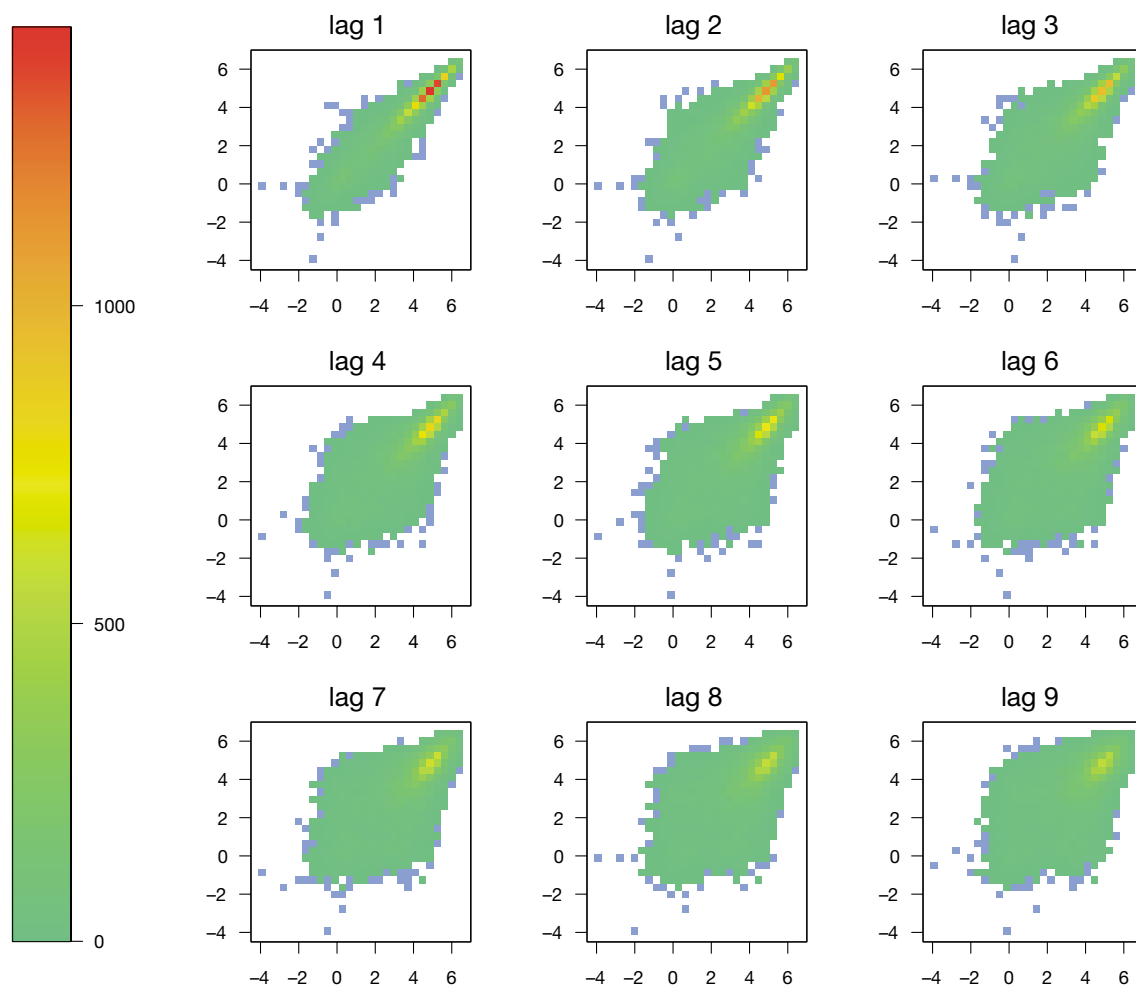


5. Plausibilisierung von Ozon und Stickoxiden



Abb. 5-2: Lag-Plots von $\log_2(\text{O}_3)$ für Schwyz

Die Skala am Rand der Grafik gibt die Anzahl Punkte pro Farbquadrat an. Einheiten der x- und y-Achsen $\log_2(\text{ppb})$.



5. Plausibilisierung von Ozon und Stickoxiden



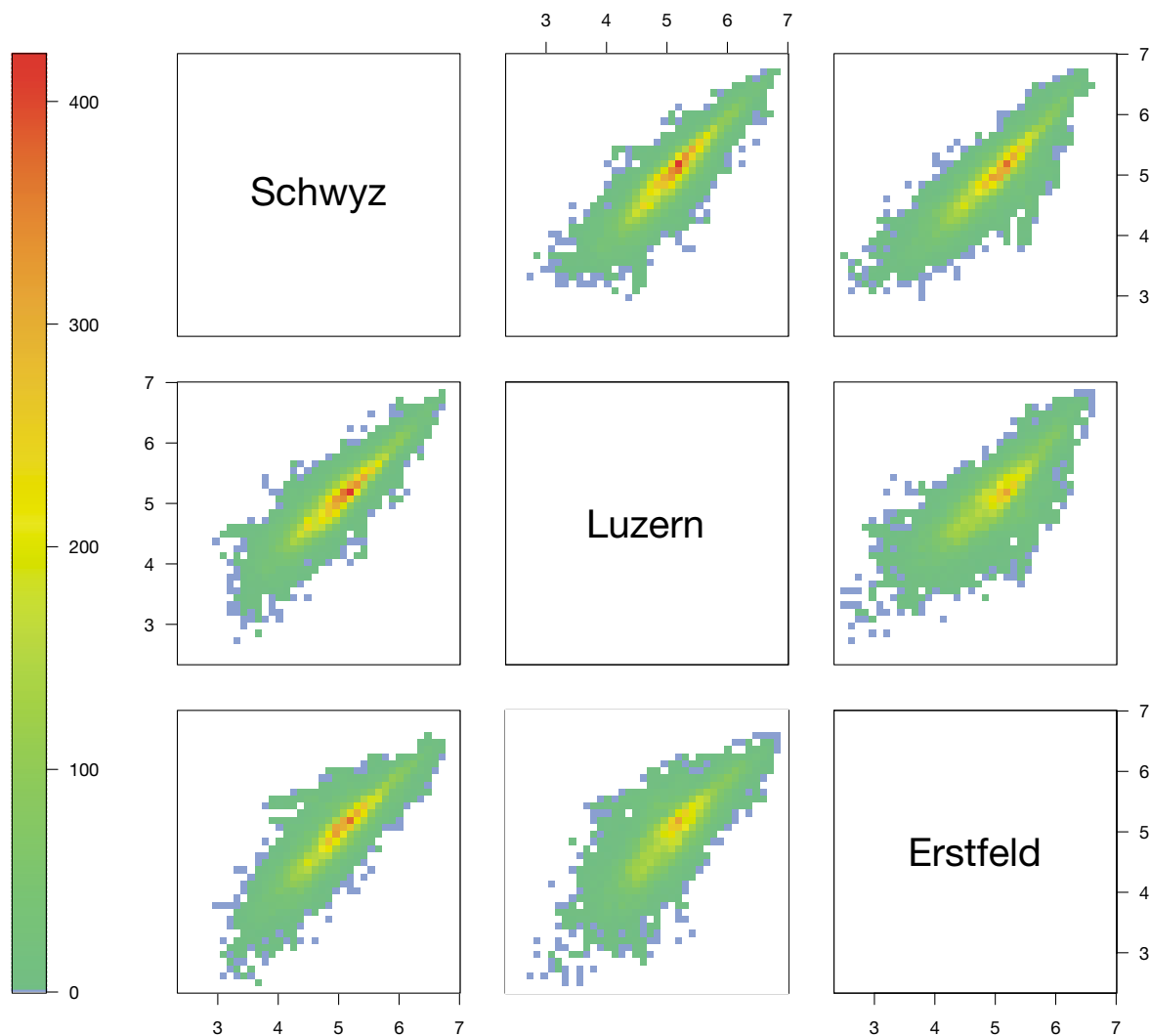
Während in den modifizierten Lag-Plots die zeitliche Korrelation der Daten studiert werden kann, dienen die modifizierten paarweisen Streudiagramme der Überprüfung der örtlichen Korrelationsstruktur. In Richtung der x-Achse werden hier die Werte der Messgrösse des Standortes 1 aufgetragen, in Richtung der y-Achse die zeitlich entsprechenden Werte des Standortes 2. Der Standort 1 des betrachteten Diagramms findet man im Schnittpunkt der Diagonalen der Diagrammmatrix und der Vertikalen durch das betrachtete Diagramm. Analog findet man den Standort 2 im Schnittpunkt der Diagonalen der Diagrammmatrix und der Horizontalen durch das betrachtete Diagramm. Die Streudiagramme in der rechten oberen Hälfte der Abbildung haben somit eine spiegelbildliche Entsprechung in der linken unteren Hälfte. In Abb. 5-3 bzw. Abb. 5-4 wurden sich die logarithmierten O_x - bzw. die O_3 -Konzentrationen gegenübergestellt, welche jeweils zur gleichen Zeit an den verschiedenen Orten gemessen wurden. Die Interpretation ist ganz analog zu derjenigen der Lag-Plots: Je stärker die Messwerte an zwei verschiedenen Orten miteinander (linear) korrelieren, um so näher liegen diese auf einer Geraden. Die örtliche Korrelation von O_x in Schwyz mit O_x in Luzern entspricht ungefähr der zeitlichen Korrelation von O_x mit einem Lag von 3 h. Dies bedeutet z. B., dass jemand am Standort Schwyz mit dem Messwert von 12.00 die Konzentration von O_x , welche um 15.00 in Schwyz herrschen wird, gleich gut voraussagen kann, wie jemand, welcher um 15.00 in der Messstation Luzern den dortigen O_x -Wert abliest und den aktuellen Wert in Schwyz schätzt. Die entsprechenden Streudiagramme von O_3 (Abb. 5-4) zeigen örtlich praktisch unkorrelierte Daten. Entsprechend zeigen auch die Lag-Plots von O_3 nur geringe Korrelation.

5. Plausibilisierung von Ozon und Stickoxiden



Abb. 5-3: Modifizierte paarweise Streudiagramme von $\log_2(\text{O}_3)$ an den Orten Schwyz, Luzern und Erstfeld

Die Skala am Rand der Grafik gibt die Anzahl Punkte pro Farbquadrat an. Einheiten der x- und y-Achsen $\log_2(\text{ppb})$.

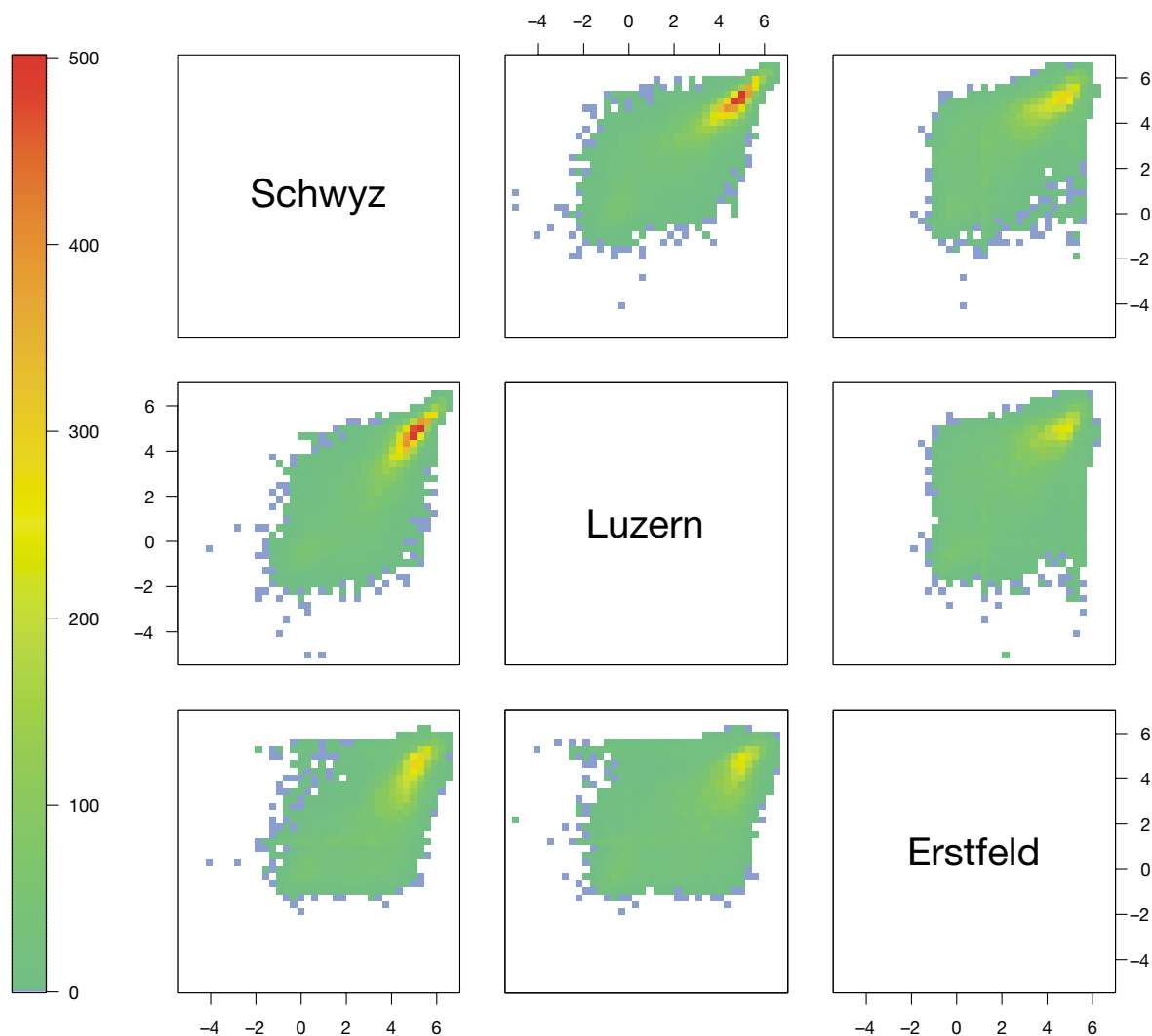


5. Plausibilisierung von Ozon und Stickoxiden



Abb. 5-4: Paarweise Streudiagramme von $\log_2(\text{O}_3)$ an den Orten Lungern, Schwyz, Luzern und Erstfeld

Die Skala am Rand der Grafik gibt die Anzahl Punkte pro Farbquadrat an. Einheiten der x- und y-Achsen $\log_2(\text{ppb})$.

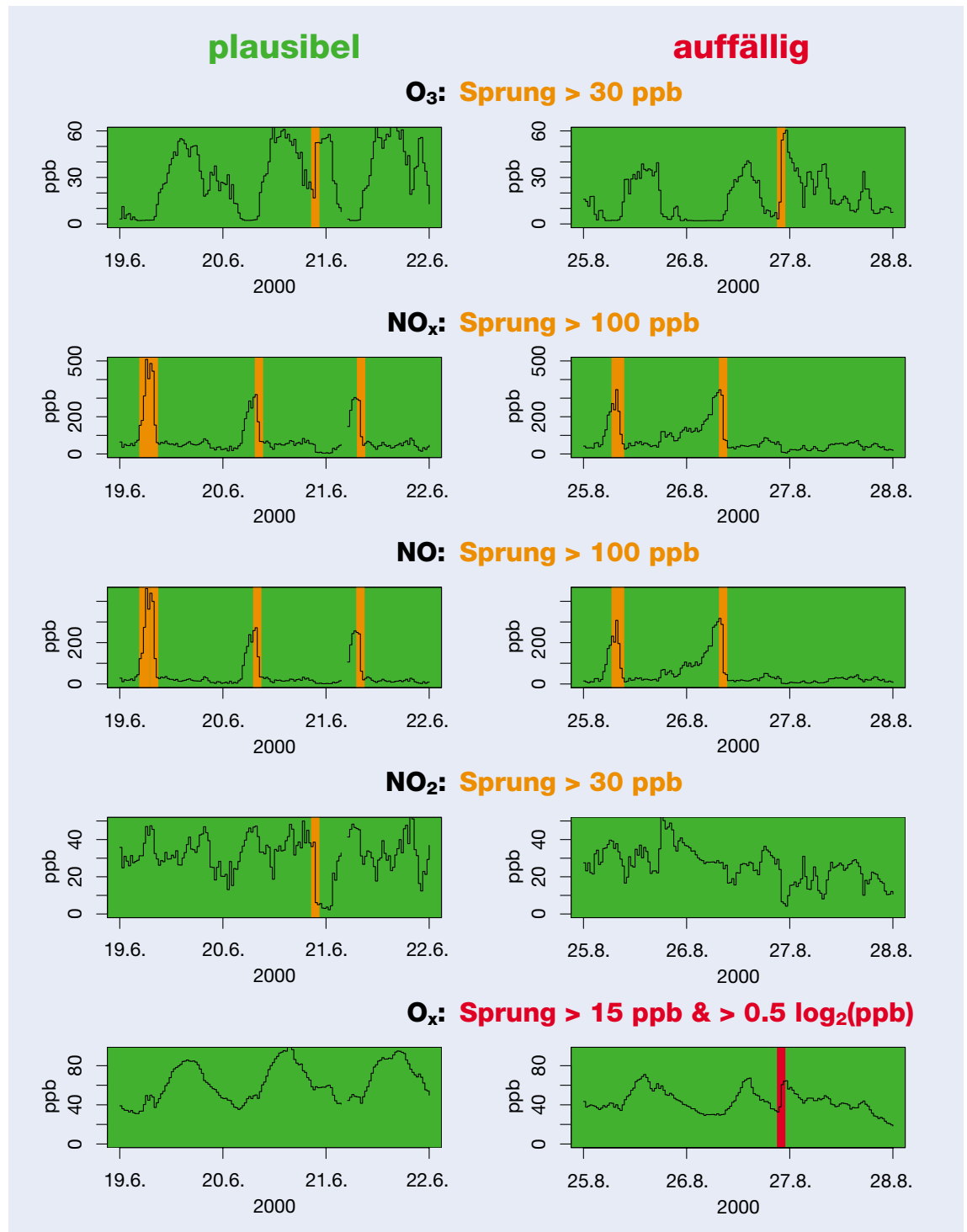


5. Plausibilisierung von Ozon und Stickoxiden



Abb. 5-5: Beispiel für Plausibilisierung via O_x in Erstfeld

In der linken Spalte sind drei Tage (19.6.2000 bis 22.6.2000) dargestellt, an denen ein grosser Sprung bei O_3 und NO_2 auftritt (orange Markierungen), welcher jedoch sehr gut mit der Titration von O_3 mit NO erklärt werden kann und somit als nicht auffällig zu gelten hat. In der rechten Spalte sind die Immissionswerte vom 25.8.2000 bis 28.8.2000 dargestellt mit einem einzigen auffälligen Zeitpunkt (rot markierter Bereich in O_x -Grafik). Der grosse Sprung im O_3 -Verlauf kann dieses Mal nur ungenügend mit der NO -Titration von O_3 erklärt werden: Föhnneffekt.



5. Plausibilisierung von Ozon und Stickoxiden



In Abb. 5-5 werden an Hand von zwei Zeitausschnitten von jeweils drei Tagen Länge nochmals die Vorteile erläutert, welche durch die Identifikation von O_x -Sprüngen entstehen an Stelle einer Überprüfung der Verläufe von O_3 , NO bzw. NO_2 . In der linken Spalte sind drei Tage in Erstfeld (19.6.2000 bis 22.6.2000) dargestellt, an denen grosse Sprünge bei O_3 , NO_x , NO und NO_2 auftreten (orange Markierungen). Für O_3 und NO_2 sind diese mit der «Titration» von O_3 mit NO erklärbar (vgl. glatter Verlauf von O_x !) und gelten somit als unauffällig. Die morgendlichen Spitzen bei NO_x und NO sind vermutlich auf den neu einsetzenden Lastwagenverkehr und das Wechseln von Berg- zu Talwind mit einer kurzen windstillen Phase dazwischen zurückzuführen. In der rechten Spalte sind die Immissionswerte vom 25.8.2000 bis 28.8.2000 vom gleichen Messort dargestellt, wo ein einziger auffälliger Zeitpunkt (roter Bereich in O_x -Grafik) markiert ist. Der als auffällig markierte O_x -Sprung, welcher auch von einem aussergewöhnlich hohen Sprung von O_3 begleitet ist (von 14.1 ppb nach 54.1 ppb innerhalb von 30 min) ist mit einem plötzlichen Föhneinfluss bei der Messstation erklärbar. Gleichzeitig mit dem Ozon stieg nämlich die Lufttemperatur von 19.9°C auf 26.8°C, sank die relative Luftfeuchte von 62 % auf 39.5 % und nahm die Windgeschwindigkeit von 1.4 m/s auf 5.2 m/s zu. Bei der Überprüfung auf ungewöhnlich grosse O_x -Sprünge werden in Erstfeld offensichtlich auch Föhneinbrüche herausgefiltert. Von insgesamt 17 solchen aussergewöhnlichen Ereignissen, konnten 9 mit Föhneffekten erklärt werden. In vier weiteren Fällen konnten die auffälligen Konzentrations-Sprünge auf zunehmenden Wind zurückgeführt werden. Sofern diese Filtereffekte nicht erwünscht sind, müsste eine Zusatzregel in den Algorithmus eingebaut werden, welche den Verlauf der Temperatur und der Windgeschwindigkeit mitberücksichtigt. An den übrigen Standorten wurden keine analogen Föhneffekte registriert.

Die grossen Sprünge bei NO_x und NO sind analog zu den morgendlichen Spitzen des 19.6. bis 22.6. erklärbar. Obwohl die Grenzwerte für die Sprünge von Ozon und Stickoxide um den Faktor 2 bis 7 höher angesetzt wurden als für O_x , wurden bei letzterem insgesamt viel weniger Ereignisse detektiert. Bei NO_x werden zwar natürlicherweise sehr grosse Sprünge erwartet, welche jedoch nicht via O_x überprüft werden können. Deshalb wurde auch für Sprünge dieser Summenkonzentration eine Begrenzung analog zu derjenigen von O_x eingeführt (vgl. Tab. 5-1). Bei beiden Summen wurde sowohl die absolute Differenz (15 ppb für O_x bzw. 60 ppb für NO_x) zwischen zwei Werten als auch die relative ($0.5 \log_2(\text{ppb})$ bzw. $2 \log_2(\text{ppb})$) begrenzt. Ein Sprung wurde dabei nur dann als auffällig taxiert, wenn sowohl die absolute Differenz als auch die relative grösser als der entsprechende Grenzwert ausfiel. Für die drei Standorte Schwyz, Luzern und Erstfeld wurden gemäss dieser Definition 6, 6 und 20 (O_x) bzw. 1, 9 und 20 (NO_x) auffällige Werte gefunden. An den gewählten Grenzen spiegelt sich wiederum das viel glattere Verhalten von O_x im Vergleich zu NO_x : Obwohl die absolute Grenze für O_x vier Mal tiefer gewählt wurde und der zulässige Faktor zwischen zwei benachbarten Werten nur ca. 1.5 statt 4 beträgt, wurden bei den drei Standorten für die Verläufe der O_x und NO_x -Konzentrationen ungefähr gleich viele auffällige Werte gefunden.

5. Plausibilisierung von Ozon und Stickoxiden



Tab. 5-1: Plausibilisierungsregeln für Ozon und Stickoxide

Falls mindestens einer der in der Tabelle aufgeführten 6 Überprüfungen positiv ausfällt oder ein Wert ausserhalb des Wertebereichs gemäss Tab. 5-2 liegt, gilt der entsprechende Zeitpunkt als auffällig. Während die Wertebereiche standortspezifisch festgelegt werden mussten, konnten hier für alle Standorte identische Parameter verwendet werden.

Überprüfung	Bedingung
Überprüfung auf «Geräteausfall».	mindestens 2 Nachkommastellen
Durchführung von unabhängigen Überprüfungen für O ₃ , NO, NO ₂ , NO _x , O _x	gleitende Standardabweichung über 3 Werte < 0.005 ppb
Definition von auffälligem O _x -Sprung	Betrag der Differenzen zweier nachfolgender O _x -Werte > 15 ppb <i>und gleichzeitig</i> Betrag der Differenzen zweier nachfolgender log ₂ -O _x -Werte > 0.5 log₂(ppb)
Definition von auffälligem NO _x -Sprung	Betrag der Differenzen zweier nachfolgender O _x -Werte > 60 ppb <i>und gleichzeitig</i> Betrag der Differenzen zweier nachfolgender log ₂ -O _x -Werte > 2 log₂(ppb)

In Tab. 5-2 sind die erlaubten und die innerhalb eines Jahres tatsächlich gemessenen Wertebereiche dargestellt. In Jahren mit aussergewöhnlichen Witterungsbedingungen wie z. B. Wintern mit lange andauernden Inversionslagen oder heissen und sonnenreichen Sommern können gehäuft Werte ausserhalb des normalen Wertebereichs auftreten und somit auch gehäuft Werte als auffällig markiert werden. Falls dies unerwünscht ist, muss der erlaubte Wertebereich entsprechend den langjährigen Erfahrungen ausgedehnt werden. Damit verringert sich aber auch automatisch die Wahrscheinlichkeit, unplausible Werte zu entdecken.

5. Plausibilisierung von Ozon und Stickoxiden



Tab. 5-2: Erlaubter und beobachteter Wertebereich für Ozon-, Stickoxid- und O_x-Halbstundenwerte in [ppb]

Beobachtungsbereich: 1.2.01 bis 31.1.02, wobei sich die Ozondaten in Lungern auf die Monate April bis September beschränken. Rot markierte Werte liegen ausserhalb des festgelegten Bereichs.

Messgrösse [Einheit]	Ort	Min. beob.	Min. erlaubt	Max. beob.	Max. erlaubt
O₃ [ppb]	Lungern	5.8	-1	96.6	100
	Schwyz	0.1	-1	94.4	100
	Luzern	0.0	-1	93.3	100
	Erstfeld	0.3	-1	73.6	80
NO_x [ppb]	Schwyz	0.1	-2	205.4	250
	Luzern	1.7	-2	413.6 ¹¹	250
	Erstfeld	0.7	-2	519.7 ¹²	500
NO [ppb]	Schwyz	-0.1	-2	182.3	200
	Luzern	-0.2	-2	354.0 ¹³	200
	Erstfeld	0.0	-2	483.2	500
NO₂ [ppb]	Schwyz	-0.4	-2	71.5	80
	Luzern	1.3	-2	71.4	80
	Erstfeld	0.0	-2	61.2	80
O_x [ppb]	Schwyz	7.6	0	107.0	120
	Luzern	6.7	0	114.2	120
	Erstfeld	5.7	0	95.2	120

Beobachtungen ausserhalb des Wertebereichs definieren zusammen mit den Regeln von Tab. 5-1 die auffälligen Zeitpunkte wie sie exemplarisch für die O_x-Zeitreihe in Luzern in Abb. 5-6 dargestellt wurden. Blau markiert sind Stellen, bei denen mindestens 3 Werte hintereinander (fast) identisch sind. Rot markiert sind Peaks, welche auffällig stark vom glatten Verlauf abweichen (O_x-Sprünge).

¹¹ 3 Werte lagen über der Limite von 250 ppb. Sie wurden an drei unterschiedlichen Vormittagen im Herbst und Winter gemessen. Gleichzeitig waren die O₃-Werte aussergewöhnlich tief und die NO-Werte hoch.

¹² 2 Werte lagen über 500 ppb. Der dritt grösste Wert betrug 472.9 ppb

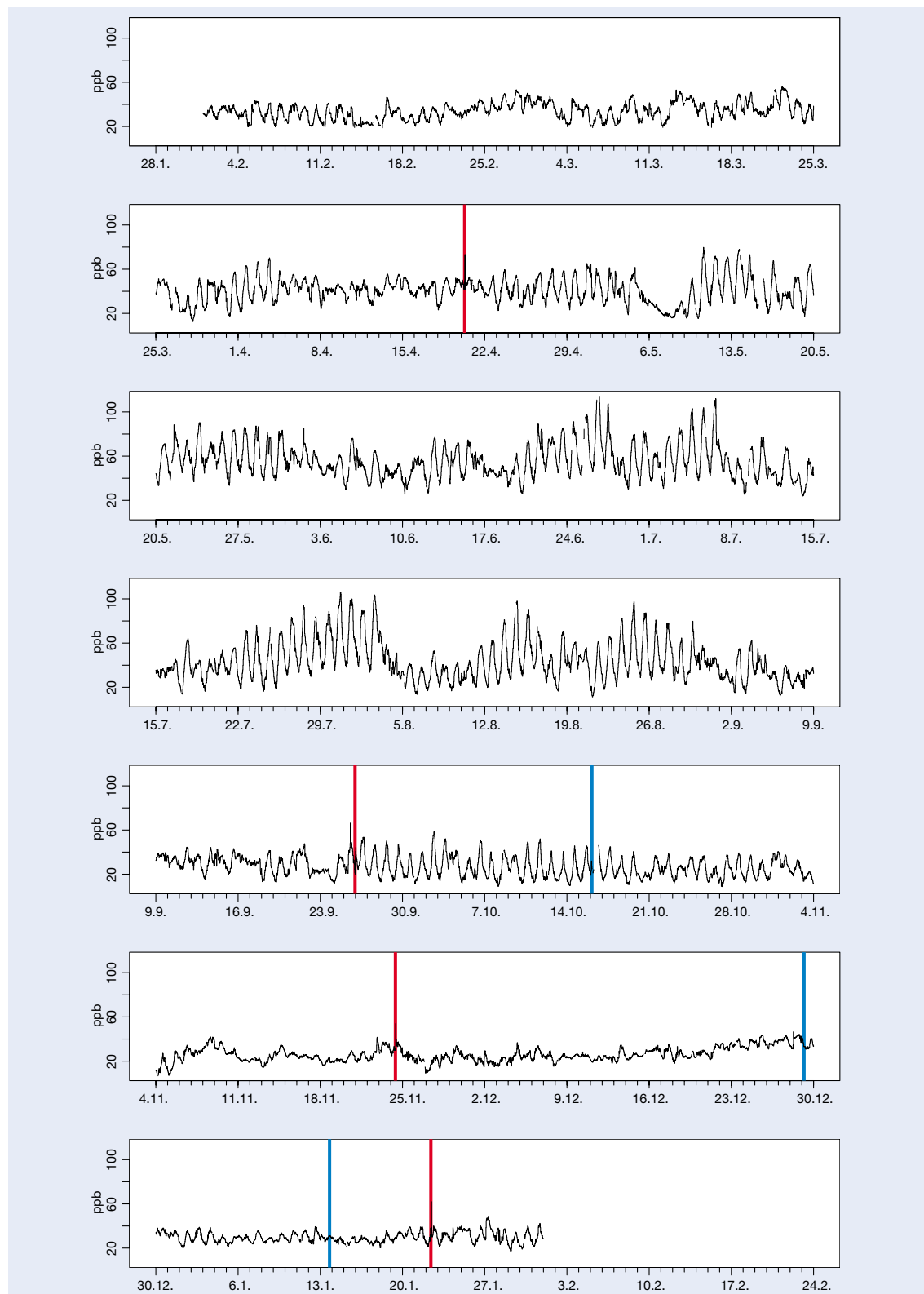
¹³ 5 Werte lagen über 200 ppb, wobei diese an 4 unterschiedlichen Tagen im Herbst und Winter auftraten.

5. Plausibilisierung von Ozon und Stickoxiden



Abb. 5-6: Zeitreihe von O_x in Luzern mit auffälligen Werten

1.2.01 bis 31.1.02. Blau markiert sind Stellen, bei denen mindestens 3 Werte hintereinander (fast) identisch sind. Rot markiert sind Peaks, welche auffällig stark vom glatten Verlauf abweichen. Auf der Zeitachse sind jeweils die Montage mit Datum versehen.



6. Plausibilisierung von PM10



In Abb. 6-1 und Abb. 6-2 wurden die bereits in Kap. 5 beschriebenen Lag-Plots bzw. paarweisen Streudiagramme erläutert. Während bei den O_x die logarithmierten Werte die beste Darstellung ergaben, zeigten bei PM10 die wurzeltransformierten Werte die besten Streudiagramme. Nähere Begründungen zu den gewählten Transformationen finden Sie in Kap. 7.2 und 7.3. Der Vergleich der Abbildungen von PM10 (Abb. 6-1 und Abb. 6-2) mit denjenigen von O_x (Abb. 5-1 und Abb. 5-3) bzw. O_3 (Abb. 5-2 und Abb. 5-4) zeigt, dass die zeitlichen und örtlichen Korrelationen für PM10 zwischen denjenigen von O_x und O_3 liegt. Es konnten keine erfolgreichen lufthygienischen Überlegungen analog zur O_x -Transformation gemacht werden, so dass hier die wurzeltransformierten PM10-Konzentrationen direkt untersucht wurden.

Tab. 6-2 zeigt die gefundenen Wertebereiche für PM10 sowie die vorgeschlagenen Bereichs- und Grenzen. Analog zu den O_x -Werten konnten für alle drei Standorte die gleichen Grenzen für die Wertebereichsgrenzen der PM10-Konzentrationen angenommen werden. Bei Schwyz und Luzern wurden je rund ein halbes Dutzend Ereignisse gefunden, welche den Bereich überschritten. Die meisten der mit der Bereichsregel gefundenen Werte wurden auch mit der Sprungregel gemäss Tab. 6-1 detektiert. Bei auffälligen Ereignissen, welche nur aus einem einzelnen Datenpunkt bestanden, ist damit mit einer erhöhten Wahrscheinlichkeit für einen tatsächlich ungültigen Wert zu rechnen. Leider existieren keine unabhängigen Beobachtungen zu diesen Zeitpunkten und der Verlauf der PM10-Konzentration korreliert auch nicht eng mit anderen Luftinhaltsstoffen, so dass über diese Zeitpunkte keine weiteren Aussagen gemacht werden können.

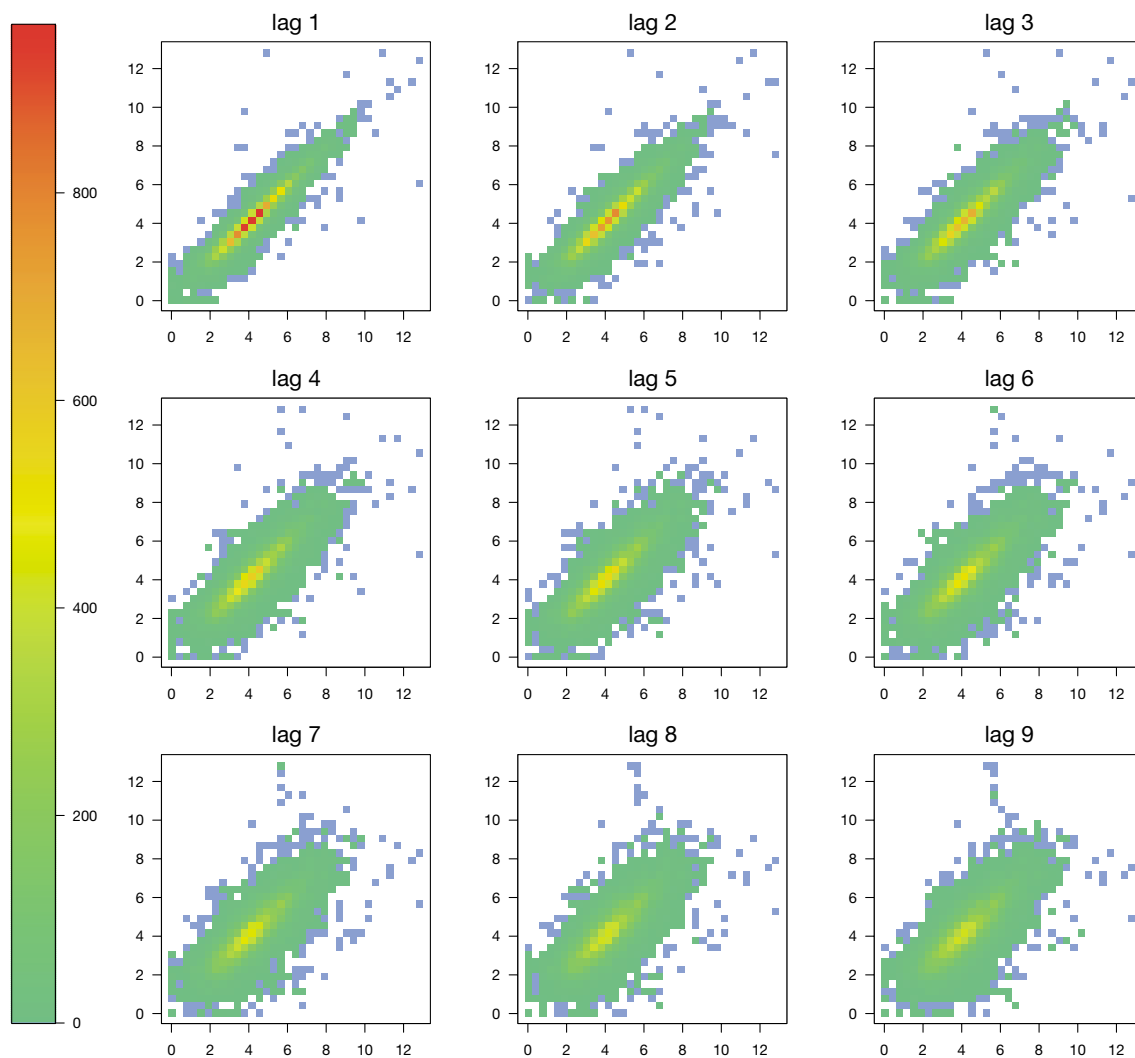
Die gemäss den Regeln nach Tab. 6-2 und Tab. 6-1 plausibilisierten PM10-Daten von Erstfeld, lieferten 4 Stellen mit nicht variierenden Werten und 9 Stellen mit aussergewöhnlich starken Konzentrationssprüngen (Abb. 6-3). Während bei O_x und NO_x aussergewöhnlich grosse Sprünge auf Grund der absoluten und relativen Differenzen (= Differenzen der logarithmierten Werte) bestimmt wurden, wurden bei PM10 die absoluten Differenzen und die Differenzen der wurzeltransformierten Grössen verwendet. Letzteres entspricht der «natürlichen» Transformation der Werte, wie sie in Kap. 7.3 erläutert wird. Der Effekt der Wurzel-Transformation ist analog zur Logarithmus-Transformation, allerdings etwas weniger ausgeprägt. Bei der Wurzel-Transformation werden absolute Differenzen im höheren Konzentrationsbereich milder beurteilt als im tiefen.

6. Plausibilisierung von PM10



Abb. 6-1: Lag-Plots von $\sqrt{\text{PM}_{10}}$ in Luzern

Die Skala am Rand der Grafik gibt die Anzahl Punkte pro Farbquadrat an.
Einheiten der x- und y-Achsen: $\sqrt{\mu\text{g}/\text{m}^3}$.

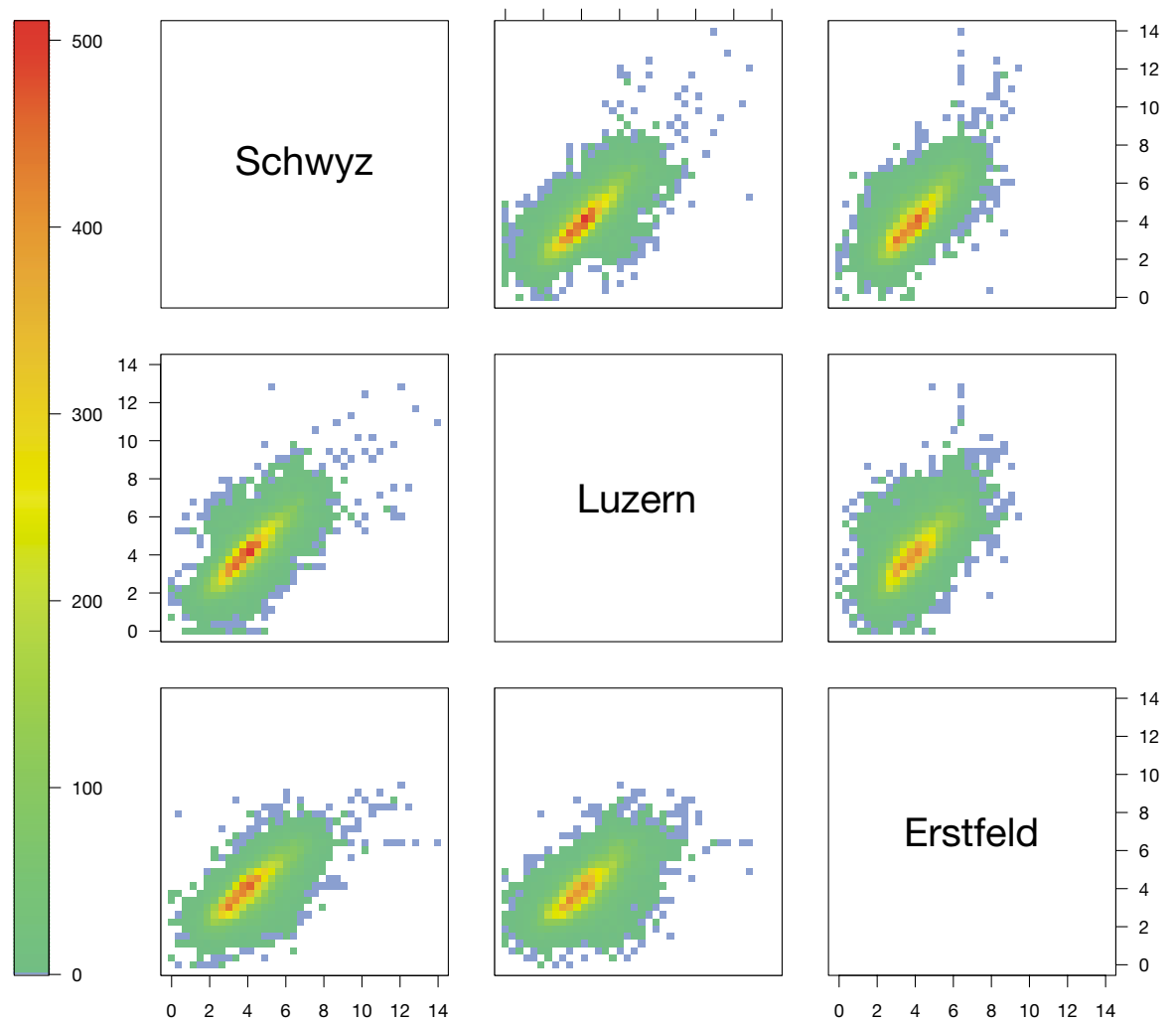


6. Plausibilisierung von PM10



Abb. 6-2: Paarweise Streudiagramme von $\sqrt{\text{PM}_{10}}$ an den Orten Schwyz, Luzern und Erstfeld

Die Skala am Rand der Grafik gibt die Anzahl Punkte pro Farbquadrat an.
Einheiten der x- und y-Achsen: $\sqrt{\mu\text{g}/\text{m}^3}$.



6. Plausibilisierung von PM10



Tab. 6-1: Plausibilisierungsregeln für PM10

Falls mindestens einer der in der Tabelle aufgeführten 6 Überprüfungen positiv ausfällt oder ein Wert ausserhalb des Wertebereichs gemäss Tab. 6-2 liegt, gilt der entsprechende Zeitpunkt als auffällig.

Überprüfung	Bedingung
Überprüfung auf «Geräteausfall».	mindestens 1 Nachkommastellen
Durchführung von unabhängigen Überprüfungen für O₃ , NO , NO₂ , NO_x , O₃	gleitende Standardabweichung über 4 Werte < 0.05 µg/m ³
Definition von auffälligem PM10-Sprung	Differenzen zweier nachfolgender Werte > 35 ppb und gleichzeitig Differenzen zweier nachfolgender √PM10-Werte > 3 √µg/m ³

Tab. 6-2: Erlaubter und beobachteter Wertebereich für PM10-Halbstundenwerte in [µg/m³]

Beobachtungsbereich: 1.2.01 bis 31.1.02.

Rot markierte Werte liegen ausserhalb des festgelegten Bereichs.

Messgrösse [Einheit]	Ort	Min. beob.	Min. erlaubt	Max. beob.	Max. erlaubt
PM10 [µg/m ³]	Schwyz	0	-2	195.4 ¹⁴	100
	Luzern	0	-2	165.7 ¹⁵	100
	Erstfeld	0	-2	91.3	100

¹⁴ Bei 6 Ereignissen mit insgesamt 19 Werten wurden Konzentrationen über 100 µg/m³ gemessen.

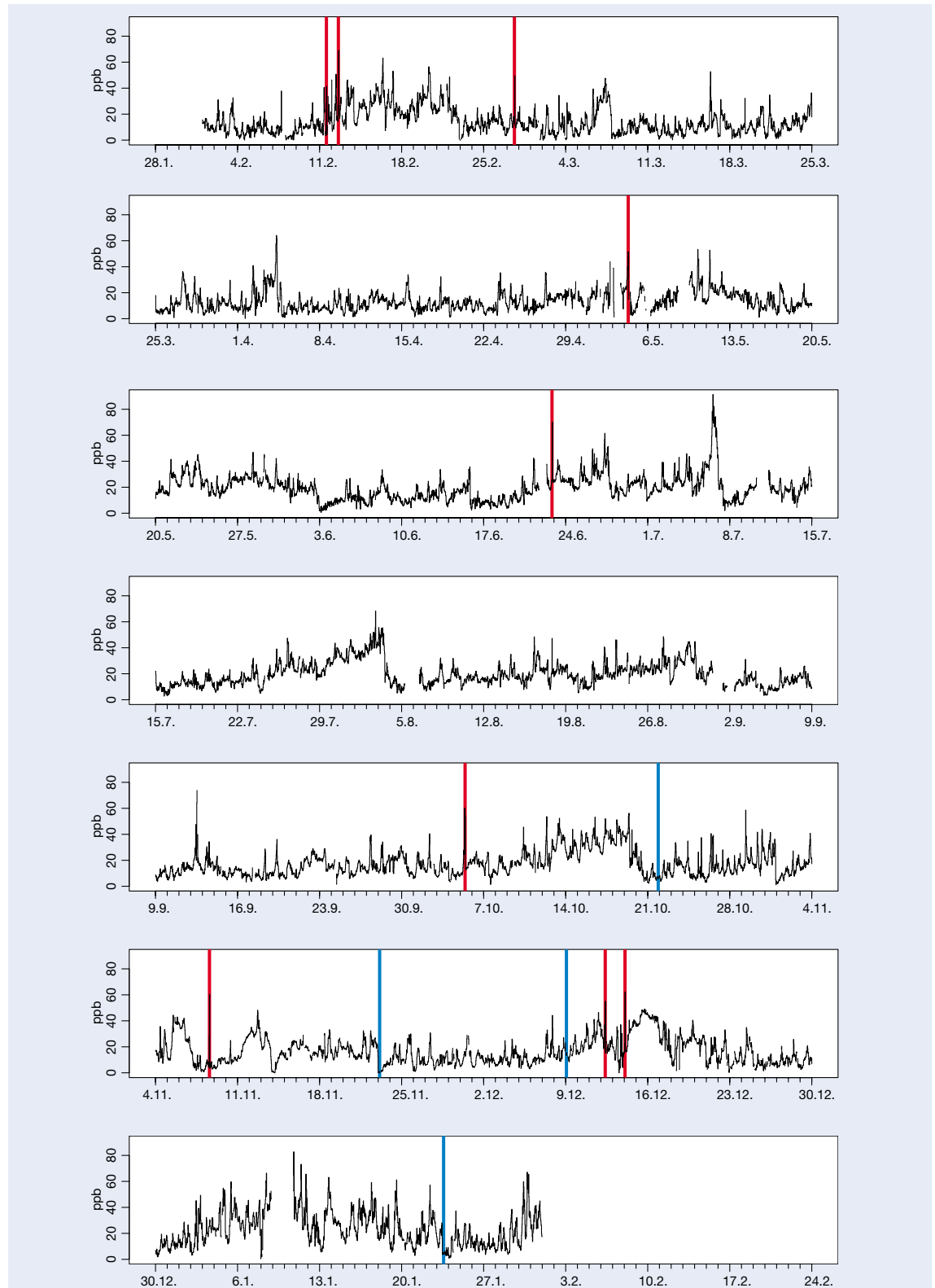
¹⁵ Bei 4 Ereignissen mit insgesamt 10 Werten wurden Konzentrationen über 100 µg/m³ gemessen.

6. Plausibilisierung von PM10



Abb. 6-3: Zeitreihe von PM10 in Erstfeld mit auffälligen Werten

1.2.01 bis 31.1.02. Blau markiert sind Stellen, bei denen mindestens 4 Werte hintereinander (fast) identisch sind. Rot markiert sind Peaks, welche auffällig stark vom glatten Verlauf abweichen. Auf der Zeitachse sind jeweils die Montage mit Datum versehen.



7. Physikalisch-chemische Interpretationen



7.1 Charakterisierung der Standorte

Der Median und der MAD sind die robusten Kennzahlen für die «mittlere» Höhe der Konzentrationen und deren Streuung (vgl. Glossar). Sie werden hier zum Vergleich der vier luft-hygienisch unterschiedlichen Standorte herangezogen.

Der Median und MAD der Halbstundenmittelwerte für die Schadstoffe NO_x , NO und NO_2 steigen von Schwyz über Luzern bis Erstfeld. Analoges gilt auch für den MAD der einfachen Differenzen zwischen zwei nachfolgenden Halbstundenwerten, welcher ein Mass für die kurzfristigen Schwankungen ist. Sowohl die «mittlere» Belastung als auch die kurz- und langfristige Variabilität der Stickoxid-Konzentrationen nehmen also in der gleichen Reihenfolge der Standorte zu.

Bei der «mittleren» Belastung mit dem Sekundärschadstoff O_3 ist die Reihenfolge auf Grund der hohen negativen Korrelation¹⁶ der O_3 -Konzentration mit NO_x gerade umgekehrt: Schwyz weist nach Lungern-Schönbühl die höchste, Erstfeld die geringste «mittlere» Konzentration auf. Die hohen Konzentrationen in Lungern sind einerseits durch die Lage («Reinluftgebiet») und andererseits durch die auf die Ozonsaison (April bis September) eingeschränkte Messaktivität zurückzuführen.

Die NO_2 -Konzentration nimmt mit steigenden NO-Emissionen bzw. NO_x -Konzentrationen zu, allerdings nicht linear. Bei hohen NO-Konzentrationen ist das Ozon, welches zur Oxidation von NO zu NO_2 gebraucht wird, limitierend. Entsprechend nimmt die Median-Konzentration von NO_2 von Schwyz nach Luzern markant zu, ist jedoch in Erstfeld nur noch minimal höher als in Luzern, obwohl die entsprechenden NO_x -Werte in Erstfeld deutlich höher liegen als in Luzern.

Für O_x weist die mittlere Konzentration und deren Variabilität nur geringe Unterschiede zwischen den Standorten auf, was die örtliche und zeitliche Glattheit der Summe der Konzentrationen von NO_2 und O_3 widerspiegelt (vgl. Kap. 5).

Bei den Konzentrationen von PM_{10} wird das Muster der Reihenfolge auf den Kopf gestellt. Hier ist nicht Schwyz der am wenigsten belastete Standort sondern Erstfeld! Allerdings sind sowohl die Belastungsunterschiede als auch die Unterschiede in den Variabilitäten zwischen den drei Standorten gering. PM_{10} scheint also analog zu O_x örtlich relativ glatt zu sein.

¹⁶ Wie in Kap. 5 gezeigt, reagiert O_3 mit NO sehr rasch zu NO_2 und wird damit abgebaut.

7. Physikalisch-chemische Interpretationen



Tab. 7-1: Vergleich der Standorte

In den Spalten sind der Reihe nach der Median und der MAD der Halbstundenwerte sowie der Median der Differenzen von je zwei aufeinanderfolgenden Halbstundenmittelwerten in [ppb] (gasförmige Stoffe) bzw. $\mu\text{g}/\text{m}^3$ (PM10) aufgelistet.

	Standort	Median	MAD	MAD(dif)
O₃	Lungern	47.1	70.1	1.7
	Schwyz	18.9	28.0	2.6
	Luzern	14.3	21.2	2.6
	Erstfeld	9.8	14.6	2.2
NO_x	Schwyz	15.8	23.4	4.1
	Luzern	26.5	39.3	6.3
	Erstfeld	43.9	65.0	10.7
NO	Schwyz	3.8	5.6	2.9
	Luzern	7.4	11.0	3.2
	Erstfeld	23.0	34.2	7.8
NO₂	Schwyz	11.5	17.0	2.0
	Luzern	17.8	26.3	2.5
	Erstfeld	18.8	27.9	2.5
O_x	Schwyz	31.7	46.9	1.7
	Luzern	33.7	50.0	1.7
	Erstfeld	32.4	48.1	1.7
PM10	Schwyz	16.5	24.5	2.5
	Luzern	17.2	25.5	2.2
	Erstfeld	15.4	22.9	2.1

7.2 Verteilung der Konzentration gasförmiger Schadstoffe

Im Bericht «VOC-Immissionsmessungen in den Kantonen Zürich, Schaffhausen und Luzern 1993–98», p. 65ff [2] wurde für die meisten Luftinhaltsstoffe, welche mit genügender Genauigkeit gemessen werden konnten, eine Lognormal-Verteilung gefunden. Ausgenommen von dieser Verteilung waren O₃, NO und NO₂. NO_x hingegen zeigte ebenfalls eine Lognormal-Verteilung. Die Ausnahmen waren diejenigen Stoffe, deren Konzentrationen sehr stark durch eine einzige schnelle chemische Reaktion beeinflusst waren. Damit war eine wesentliche Voraussetzung (viele kleine Effekte beeinflussen multiplikativ die Konzentration, vgl. Kap. 3.1) für eine Lognormal-Verteilung nicht mehr gegeben. Die Summenkonzentration O_x = O₃ + NO₂ erfüllt hingegen die Voraussetzungen, analog zur Summenkonzentration NO_x = NO + NO₂. In beiden Fällen wird durch die Summation der Effekt der schnellen und dominierenden chemischen Reaktion eliminiert. Tendenziell ist die Anpassung an die Lognormal-Verteilung für O_x schlechter als für NO_x. Besonders stark sind die Abweichungen in Erstfeld, wobei dies für die vorliegende Arbeit von untergeordneter Bedeutung war.



7.3 Verteilung der Partikelkonzentrationen

Bei PM₁₀ handelt es sich um Partikel, welche um Grössenordnungen grösser sind als einzelne Gasmoleküle. Falls die Form, Grösse und Zusammensetzung der Partikel sich wenig ändert, wird eine Poisson-Statistik (vgl. Kap. 3.1) für diese Messgrösse erwartet. Die hier betrachteten Differenzen werden zwischen zwei Werten berechnet, welche unmittelbar nacheinander, innerhalb von 60 min gemessen wurden. Die Variabilität der PM₁₀-Werte wird jedoch kurzzeitig (Minuten-Skala) in erster Linie durch die Variabilität der Quellenaktivitäten¹⁷ verursacht. Die Merkmale der Partikel ändern hingegen in der Regel auf einer Skala von einigen Stunden oder Tagen (Änderungen der meteorologischen Bedingungen, Auftreten von Saharastaub, Pollenflug etc.). Poissonverteilte Werte haben nach einer Wurzel-Transformation konstante Varianz. Dies ist bei PM₁₀ tatsächlich relativ gut der Fall, sofern man die Nullwerte, welche auf Artefakte der Messtechnik zurückzuführen sind, vernachlässigt.

Die Differenzen der PM₁₀-Werte sind im Unterschied zu den andern untersuchten Grössen auch in den Extremwerten symmetrisch. Dies bedeutet, dass sowohl der Aufbau der Konzentrationen in der Luft wie auch deren Abbau bzw. Auswaschung – mindestens bei der Beobachtung der Konzentrationen mit einer Zeitauflösung von 30 min – gleich schnell erfolgt. Besonders extreme Auswaschungseffekte bei Partikeln können also nicht häufiger festgestellt werden als besonders schnelle Anstiege der Partikelkonzentrationen.

¹⁷ z. B. Änderung der Anzahl an der Messstation passierender Lastwagen.

8. Literatur



- [1] Empfehlungen über die Immissionsmessung von Luftfremdstoffen; BUWAL; Bern 1990
- [2] VOC-Immissionsmessungen in den Kantonen Zürich, Schaffhausen und Luzern 1993–1998; René Locher; Umweltmaterialien Nr. 118; BUWAL; Bern 1999

9. Glossar



Hier werden spezielle Begriffe erklärt, welche in diesem Bericht verwendet werden. In der rechten Spalte kursiv gedruckte Begriffe verweisen auf einen weiteren Eintrag im Glossar.

Begriff	Erklärung
auffällig	Auffällige Daten unterscheiden sich von unauffälligen in Bezug auf eine oder mehrere überprüfbare Regeln. Die Begriffe « <i>ungültig</i> » und « <i>auffällig</i> » sind nicht deckungsgleich (vgl. Kap. 2).
Differenzenbildung, einfache	Aufeinander folgende Werte einer <i>Zeitreihe</i> werden voneinander subtrahiert und bilden somit eine neue Zeitreihe. Die Differenzen können als <i>Residuen</i> der einfachsten möglichen <i>lokalen Regression</i> betrachtet werden: Nachfolgender Wert = vorhergehender Wert.
Erwartungswert	Wert, welcher dem Modell als theoretischer Wert zur Beschreibung der Lokation zu Grunde liegt.
glatt	Eine Zeitreihe heisst glatt, wenn die aufeinander folgenden Werte nicht sprunghaft sondern allmählich ändern.
kritisch	Als kritisch werden Werte bezeichnet, welche auffällig sind und durch die Fachperson «gefühlsmässig» als sehr zweifelhaft beurteilt werden. Trotzdem findet die Fachperson kein objektives Argument, um den betreffenden Wert definitiv als ungültig zu erklären.
Lag-Plot	Lag-Plots werden zur Analyse der Korrelationsstruktur von Zeitreihen verwendet: In Richtung der x-Achse werden die Werte der zu analysierenden Zeitreihe aufgetragen. In Richtung der y-Achse werden die zu den x-Werten um 1, 2, 3 usw. Zeiteinheiten verschobenen x-Werte aufgetragen. Ein ausführliches Beispiel finden Sie auf S. 14
lognormale Verteilung	Die logarithmierten Werte sind normalverteilt.
MAD	Median der absoluten Abweichungen vom <i>Median</i> . Ist (nach Multiplikation mit 1.4828) eine <i>robuste Schätzung der Standardabweichung</i> einer normal-verteilten Grösse.
Median	Die Hälfte der betrachteten Werte sind kleiner oder gleich diesem Wert.
Mittel, gewichtet	$\bar{X} = \frac{\sum_{i=1}^n w_i X_i}{\sum_{i=1}^n w_i}$ <p>Für $w_i = 1$ erhalten wir das gewöhnliche Mittel.</p>
Modell	Mathematische Formulierung der Zusammenhänge zwischen verschiedenen Variablen (Grössen). Die Zusammenhänge können physikalisch-chemische Hintergründe haben, können aber auch rein statistische Beschreibungen der Daten wiedergeben.
plausibel	Plausibel ist ein Wert, der keine Auffälligkeiten zeigt oder deren Auffälligkeit durch eine natürliche Ursache erklärt werden kann.



Begriff	Erklärung
Regression, lokal	Eine lokale Regression wird z. B. zur Vorhersage eines Wertes in einer Zeitreihe verwendet. Für die Regression werden in diesem Fall nur die unmittelbar vor dem vorherzusagenden Wert gemessenen Werte verwendet. Allgemein werden für eine lokale Regression alle in der Nähe des interessierenden Punktes liegenden Werte verwendet, wobei die Gewichtung der Punkte mit zunehmendem Abstand vom Zentrum abnimmt.
Residuum	R_i Abweichung des gemessenen Wertes vom geschätzten. Bei <i>Zeitreihen</i> wird im einfachsten <i>Modell</i> die Schätzung als <i>gewichtetes Mittel</i> aus den früher gemessenen Werten berechnet, wobei das Gewicht w_i der eingehenden Messwerte (und somit der Einfluss dieser Werte auf den geschätzten Wert) mit zunehmender zeitlicher Distanz abnimmt. Falls in der Zeitreihe (lokale) Trends vorhanden sind (z. B. Konzentrationsschwankungen als Funktion der Tageszeit) wird für den aufsteigenden bzw. absteigenden Teil besser eine <i>lokale Regression</i> verwendet.
Schätzung der Standardabweichung, robuste	$1.4828 \cdot \text{MAD} (x_i - x_{\text{Median}})$
Schätzung, robuste	Statistischer Fachausdruck für die Quantifizierung eines Parameters, bei der einzelne Werte keinen übermässigen Einfluss haben.
Streudiagramme, paarweise	In Richtung der x-Achse werden die Werte der Grösse 1 aufgetragen, in Richtung der y-Achse die (z. B. zeitlich) entsprechenden Werte der Grösse 2. Ein ausführliches Beispiel finden Sie auf S. 16.
Streuung, empirische Standardabweichung	$s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}} = \sqrt{\frac{\sum_{i=1}^n (R_i)^2}{n-1}}$
ungültig	Als ungültige Daten werden Werte bezeichnet, welche mit grosser Sicherheit nicht den zur Messzeit herrschenden «wahren» Wert wiedergeben. Die Begriffe «ungültig» und «auffällig» sind nicht deckungsgleich (vgl. Kap. 2).
Varianz	Quadrat der Standardabweichung
x_i, y_i	Werte zur Zeit t_i
Zeiteinheit	Bezieht sich auf das für die Messwerte relevante Zeitintervall (vgl. Zeitfenster in eta). Für Halbstundenmittelwerte beträgt die Zeiteinheit z. B. 30 min. Hier werden nur konstante Zeitintervalle betrachtet.
Zeitreihe	Zeitlich aufeinanderfolgende Werte. In den hier behandelten Fällen handelt es sich um Zeitreihen mit konstanten Zeitintervallen, bei denen allerdings Werte fehlen können.